Stochastic models for the spread of infection

Martina Morris
Departments of Sociology and Statistics
University of Washington

Sociobehavioral and Prevention Research Core
Scientific Program on Mathematical Modeling
UW CFAR

sto·chas·tic **(st -k s t k)**

adj.

1. Of, relating to, or characterized by conjecture; conjectural.

2. *Statistics*

a. Involving or containing a random variable or variables: *stochastic calculus.*

b. Involving chance or probability: *a stochastic stimulation.*

[Greek stokhastikos, from stokhast s, *diviner*, from stokhazesthai, *to guess at*, from stokhos, *aim, goal*; see stegh- in Indo-European roots.]

Basic elements of the stochastic model

- System elements
  - Persons/animals, pathogens, vectors
- States
  - e.g., properties of persons
    - S, I, R or other indicators of infection status
    - Demographic attributes
    - Activity levels
    - Additional heterogeneity
- Rates
  - Movement from one state to another: ***Probabilistic***

Deterministic vs. stochastic models

Simplest example: Epidemic growth in a very large population

- States: only I is tracked, population has an infinite number of susceptibles
- Rates: only λ, the force of infection

	Deterministic	Stochastic
Incidence	$\frac{\partial I}{\partial t} = \lambda I$	$p(\Delta I_t = x \mid I_t) = p$
Prevalence	$I_t = I_0 e^{\lambda t}$	$\int_{t=0}^{T} \Delta I_t \ \partial t$

What does this stochastic model mean?

$p(\Delta I_t = x \mid I_t)$ Depends on the model you choose for *p(•)*

Example: Poisson – simple constant rate of new events

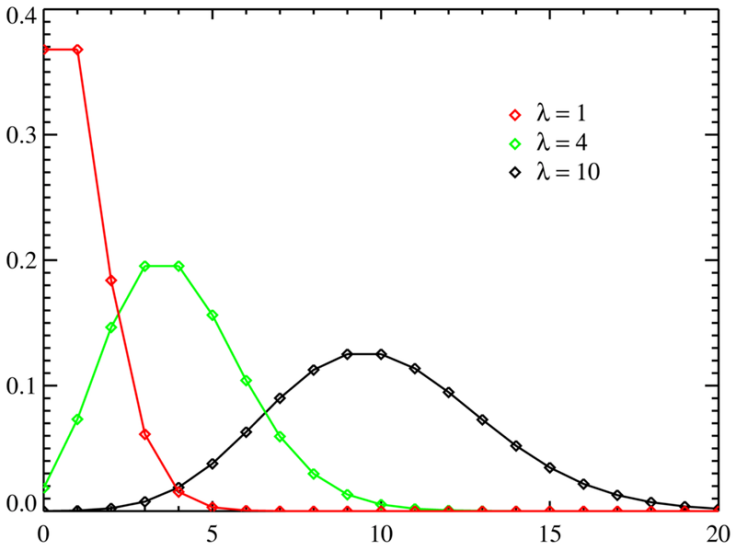
$$p(\Delta I_t = x \mid I_t) = \frac{\mu^x e^{-\mu}}{x!}$$

mean: $E(\Delta I_t) = \mu$

variance: $Var(\Delta I_t) = \mu$

$\mu = \lambda I_t \partial t$

The λ factor has this effect



note that $\frac{\mu}{\partial t} = \lambda I$ the expected number of new cases in an interval is the deterministic model prediction

Example: simple R code for this model

```
n <- 70         # the number of time steps
delta.t <- 0.01 # step size. Total time on n*delta.t
lambda <- 5     # rate per unit time
p <- rep(0,n)   # to store the prevalance
i <- rep(0,n)   # to store the incidence
p[1] <- 1       # initial prevalence

# set up a plot
plot(x=delta.t*(1:n),y=1:n, ylim=c(1,20), type="n", xlab="time", ylab="number of infections")

# step through time
for(k in 1:(n-1)){
        i[k] <- rpois(n=1,lambda=lambda*p[k]*delta.t)
        p[k+1] <- p[k]+i[k]
        points(x=delta.t*k,y=p[k+1],pch=19,cex=0.2,col=2)
}

# plot the exponential deterministic model
lines(x=delta.t*(1:n), y=exp(lambda*delta.t*(0:(n-1))), col=2)
```

This, and the rest of the code, is in the simPoisson.r file on the symposium website

Is the stochastic-deterministic relation simple?

- Will the stochastic mean ever equal the deterministic mean?
  - Yes, but only for the linear model
  - The variance of the empirical stochastic mean depends on the number of repetitions

- Can you represent variation in deterministic simulations?
  - Sensitivity analysis shows how outcomes depend on parameters
  - Parameter uncertainty can be incorporated via Bayesian methods
  - But true stochastic variation can not be represented.

- Will stochastic variation always be the same?
  - No, can specify many different distributions with the same mean
  - Process may have the same mean, but different variance
    - Negative binomial
    - Geometric
    - Many other possibilities

Focus: The influence of partnerships on epidemics

- Two fundamental questions in epidemic modeling
  - Epidemic thresholds (R_0 or not R_0, that is the question)
  - Prevalence disparities

- Will examine how partnerships affect both

Epidemic thresholds

Most basic question: can transmission be sustained in a population?

- Depends on E(*transmissions*) from the first infected case: $\mathbf{R_0}$
  - There is a threshold at $R_0 = 1$
  - Under (many) simplifying assumptions, $R_0 = \beta cD$
- The threshold means that epidemic potential is highly nonlinear
- Also means that small changes can have large impacts

The simplifying assumptions for $R_0 = \beta cD$

- **Look at the dimensional analysis:**

$$\beta cD = \frac{\text{transmission}}{\text{contact}} * \frac{\text{contacts}}{\text{time}} * \text{time}$$

  - Implies every contact is independent – i.e., no partnerships
  - Might work for vector, water, and airborne infections (malaria, cholera and flu)
  - But not for sexually transmitted infections, as contact is often with the same person

- **Can we represent partnerships in the expression for R_0?**

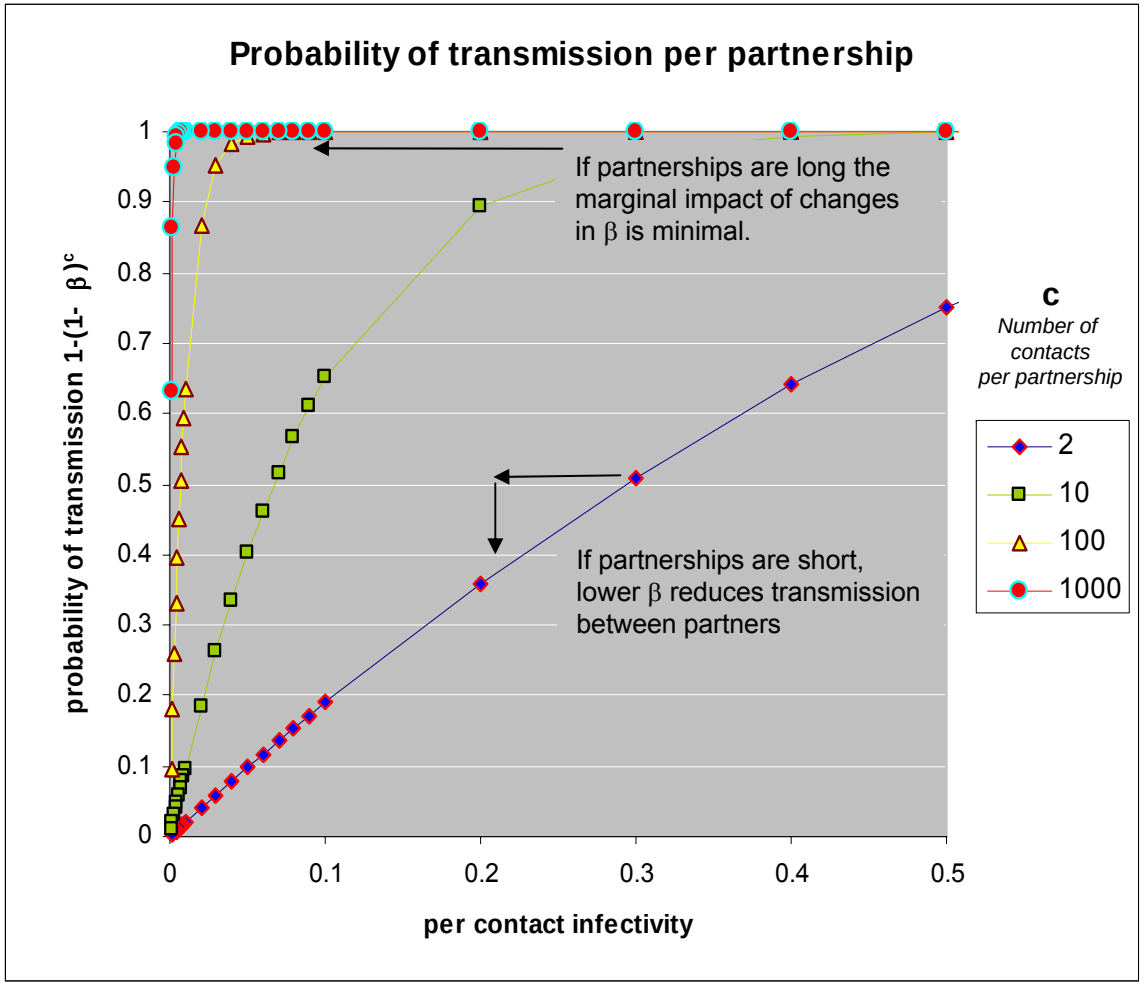
$$\tau pD = \frac{\text{transmission}}{\text{partnership}} * \frac{\text{partnerships}}{\text{time}} * \text{time}$$

… where $\tau = 1 - (1-\beta)^c$

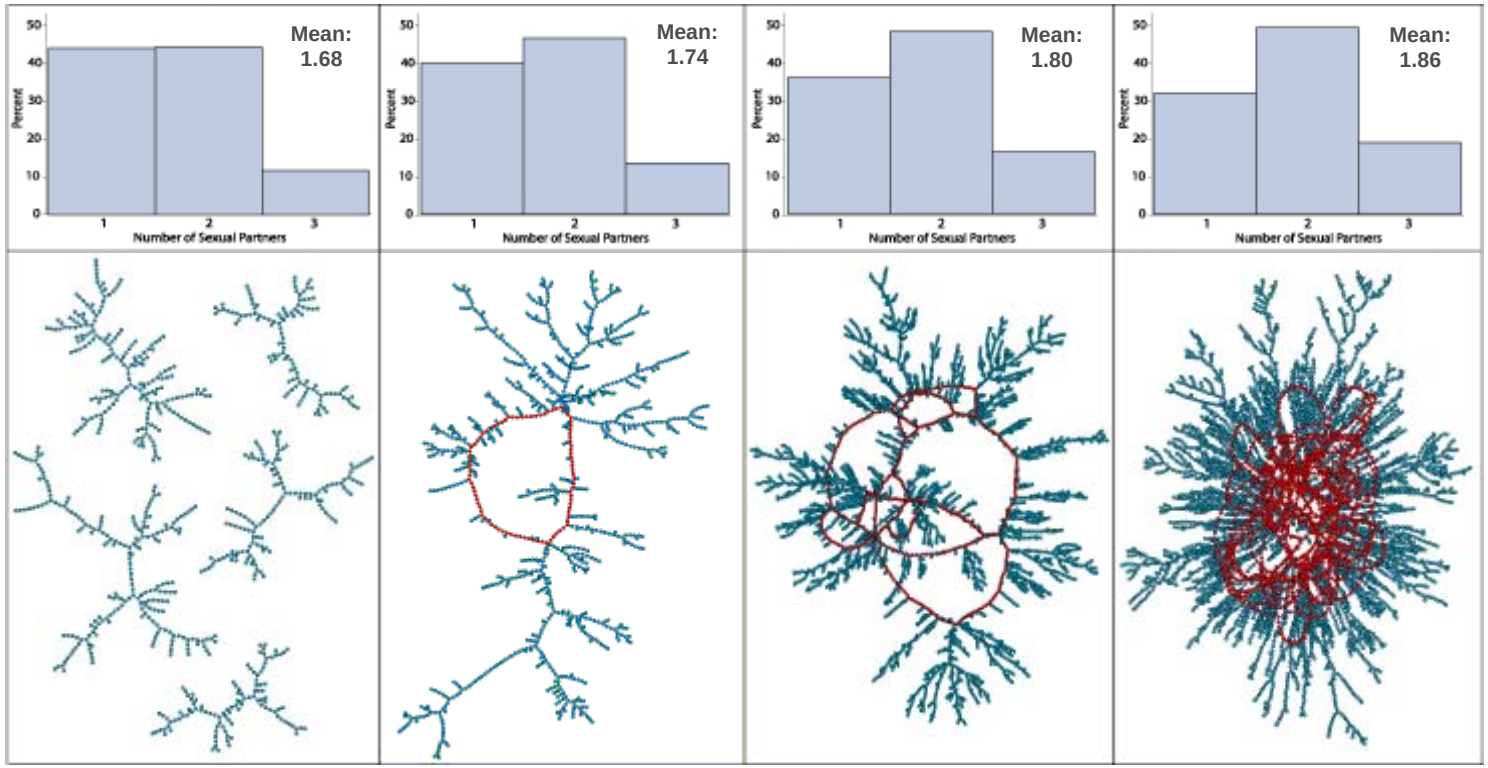
First : β influences transmission *within* partnerships

As in Susie Cassel's model, the probability of transmission within a partnership is a function of both β and c:

$1 - (1-\beta)^c$



Second : small changes in p can have threshold impacts

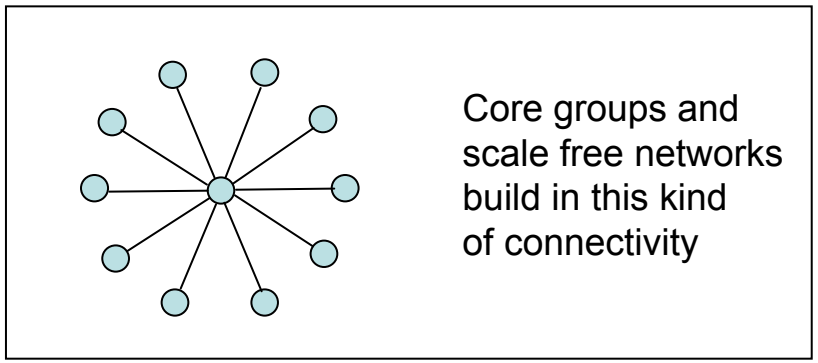


In largest component:	2%	10%	41%	64%
In largest bicomponent:	0	1%	5%	15%

Source: Morris, Goodreau and Moody 2007

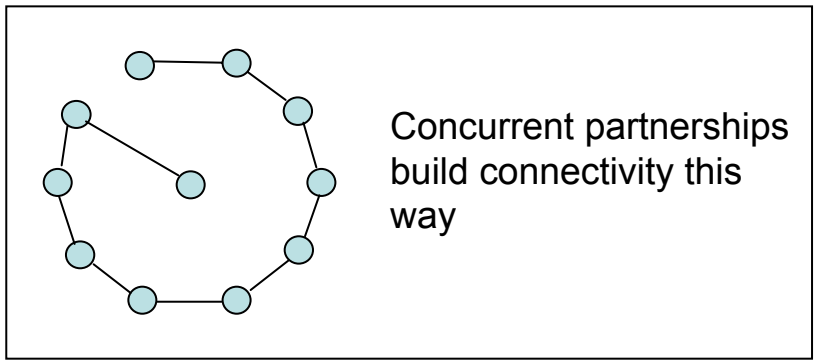
Implication: don't need core groups or "superspreaders" to create generalized epidemics

- High degree hubs



A small number of extremely active persons, with many sequential partners in a short period

- Low degree linking



Not many partners, maybe one on the side, so partnership intervals overlap, and sequence no longer protects

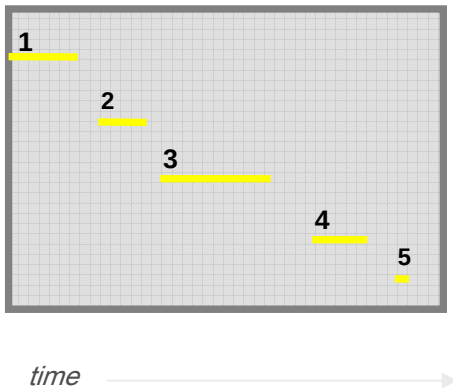
Both of these configurations have the same number of persons and partnerships

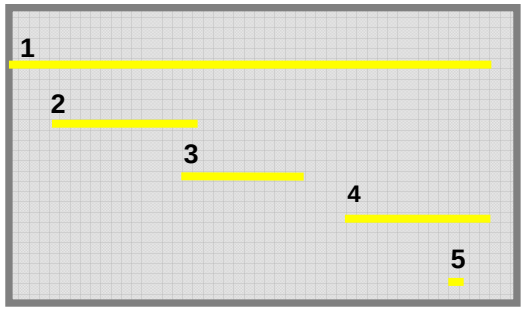
The same average contact rate, but a different distribution

Very different prevention implications (targeting, and relational context)

Third : The sequencing of partners is critical

The difference between serial monogamy and concurrent partnerships



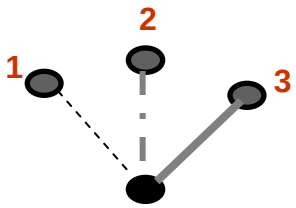


Same contact rate (5/yr), but the timing and sequence of partnerships is different

Not just a matter of reducing the interval between partnerships to 0…

Why concurrency matters

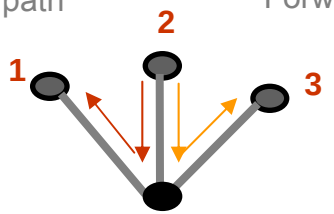
1. Removes the protection of sequence



monogamy

Backward path

Forward path

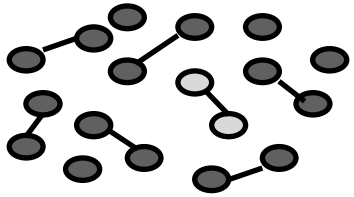


concurrency

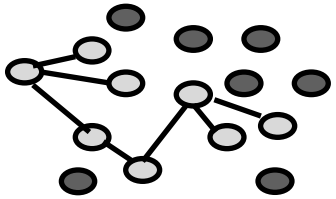
Backward path: New chain of infection

Forward path: Less time lost locked in partnership

2. Larger components in the network



monogamy



concurrency

… this is why the woman with one partner gets infected

Implication:
Concurrency may explain disparities in HIV/STI

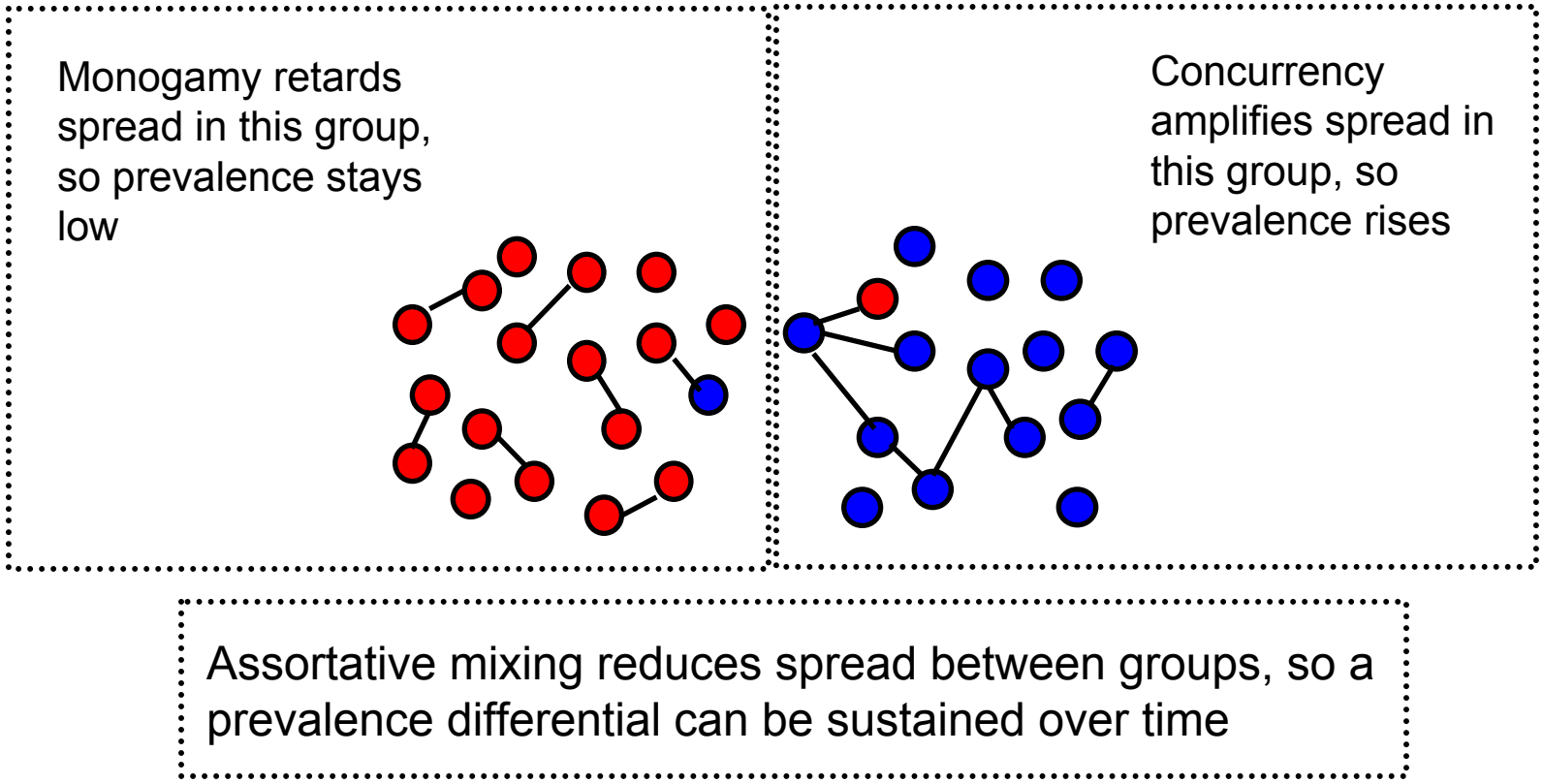
% of Men reporting concurrent partners

Observation interval	Uganda	US: NHB	US: Other races
Cross-section	14.1	8.5	3.5
In last 1mo	15.2	15.3	5.6
In last 2mo	20.0	20.2	8.0
In last 6mo	21.4	22.7	10.0
Year	26.9	25.2	12.2

*NHB Non-Hispanic Black

US NHB concurrency is similar to Uganda, though a bit lower in the cross-section

Network hypothesis for persistent disparities in HIV/STI:



Examples: Sub-Saharan Africa vs. other countries, Racial disparities in US

Example: Explaining racial disparities in HIV/STI in US

- Stochastic model for network dynamics
  - Tie formation
  - Tie dissolution
- Data from National Longitudinal Survey of Adolescent Health
- Simulation of epidemic potential
  - Comparing concurrency pattern observed in the data
  - To a monogamous population with the same number of partnerships

Data source: Add Health

- **National Longitudinal Survey of Adolescent Health**

  Cohort of 20,000 respondents, with 3 waves of data

  Wave 3 from 2000-1 contains data on 18-25 year olds:

Biomarkers	Local sexual network data
Gonorrhea Trichomoniasis Chlamydia HIV	Partners in last 5 years Age, race, sex of each partner Dates of 1st & last sex (duration) Current status of partnerships

Add Health STI prevalence ratios: non-Hispanic Black to White

STI:	***Prevalence Ratio***	
	Add Health	***CDC***
Chlamydia ***(Miller et al., STD 2005)***	***6.5***	***6.5***
Gonorrhea ***(Miller et al., JAMA 2004)***	***21.8***	***23.6***
Trich. ***(Miller et al., JAMA 2004)***	***5.9***	*****
HIV ***(Morris et al., AJPH 2006)***	***22.4***	***20***

Add Health Data: Concurrency

These are the “momentary degree distributions” that measure concurrency.

About 4% of whites report concurrent partners, with little difference between sexes

About 12% of black male and 7% of black females do.

The means for non-isolates are:

1.06, 1.08 for WF WM
1.12, 1.26 for BF BM

		White	Black
male	0	2250	369
	1	2369	430
	2	125	55
	3	24	23
	4	2	2
	5	0	5
	6	0	2
	8	0	1
Total		4770	887
female	0	1715	324
	1	3101	604
	2	131	53
	3	17	12
	4	2	0
	7	2	0
Total		4968	993

Add Health Data: Mixing patterns by race

As with US adults, mixing is strongly assortative.

Men's reports	White	Black	Other
White	85.4	1.7	12.9
Black	12.7	77.5	9.7
Other	26.7	4.3	69.0

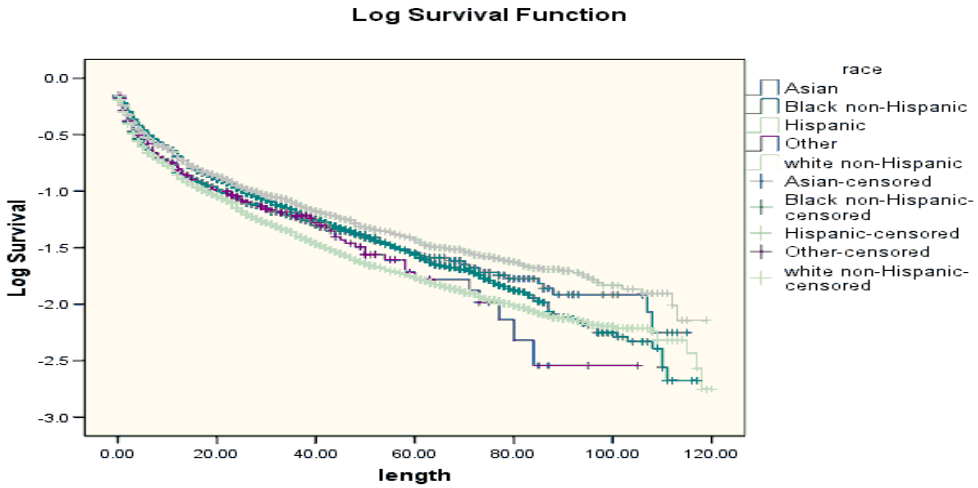
Women's reports			
White	81.7	5.4	12.9
Black	3.3	89.6	7.1
Other	19.0	8.6	72.4

Add Health Data: Partnership length

Means and Medians for Survival Time

race	Mean[a]				Median			
			95% Confidence Interval				95% Confidence Interval	
	Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
white non-Hispanic	27.040	.373	26.309	27.771	8.000	.220	7.569	8.431
Black non-Hispanic	30.878	.611	29.681	32.076	12.000	.435	11.147	12.853
Hispanic	34.542	.817	32.940	36.145	12.000	.613	10.798	13.202
Asian	30.410	1.228	28.004	32.815	8.000	.723	6.583	9.417
Other	26.400	2.239	22.011	30.789	8.000	1.537	4.988	11.012
Overall	29.078	.291	28.508	29.648	9.000	.170	8.667	9.333

a. Estimation is limited to the largest survival time if it is censored.



Stochastic model for partnership network (ERGM)

Focus: modeling the probability of forming a partnership

$$\text{logit}(p(x_{ij}=1 \mid \mathbf{X}^c)) = \theta + \theta_1 \delta(\mathbf{x}_1) + \theta_2 \delta(\mathbf{x}_2) + \ldots + \theta_n \delta(\mathbf{x}_n)$$

where: $\delta(x)$ = vector of network change statistics

θ = vector of model parameters

A "change statistic" is the change in the count of a network configuration if a specific dyad state xij is toggled {0 to 1, or 1 to 0}. For example:

- Number of ties
- Number of nodes with degree 2
- Number of partnerships between persons in the same demographic group

A stochastic simulation to show the impact of concurrency

- Primary focus: **the reachable path of infection**
- We get this by setting $\beta = 1$.
  - Guarantees all partners of infected nodes will be infected
  - So all transmission is determined by the contact network, partnership duration and sequence
- 10,000 node network, with 2 races, and 10 initial infection seeds
- Simulated over 10 years, no vital dynamics
- Match all of the target Add Health statistics
- Compare to a monogamous population with the same # of partnerships

Basic simulation code

Setup

- Estimate ERGM model for partnership network on Add Health
- Simulate a network with the ERGM estimates
- Seed 10 nodes with infection

Dynamic simulation:

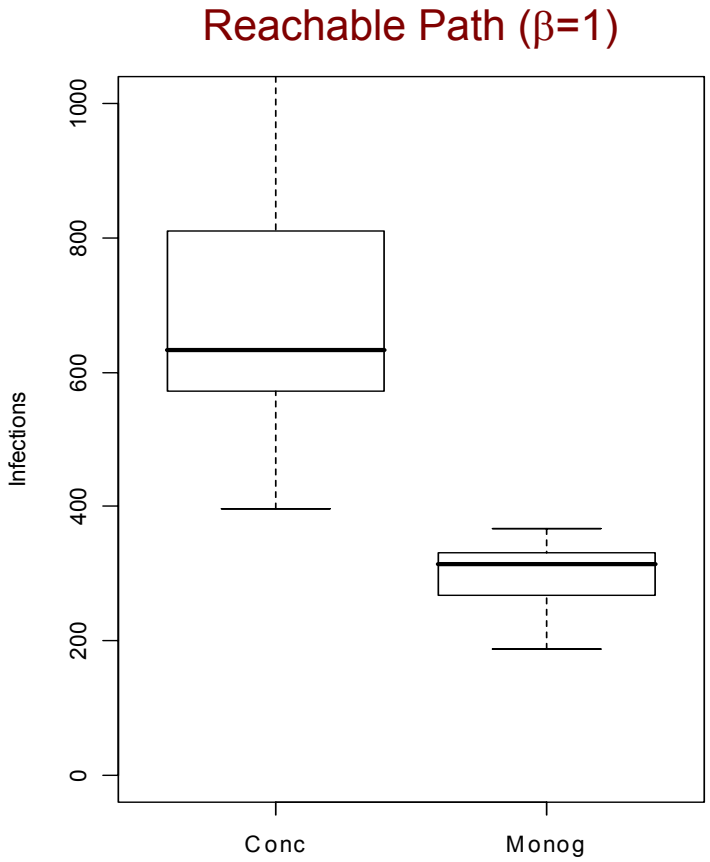
- Transmission step
  - For each discordant pair, transmit infection (since β=1)
- Pair dissolution step
  - Evaluate all pairs, dissolve with probability = 1/Duration
- Pair formation step
  - Draw two nodes, form pair with probability governed by ERGM
  - Repeat until target number of pairs is reached
- Repeat

Basic results

After 10 years (and over many different simulation runs)

- **The mean number of cumulative partners for this population is 3-4.**
  - 94% have 6 or fewer
  - The maximum observed is 12-15
- **The cumulative connected component includes almost everyone**
  - Only 97 of the 10,000 nodes are not connected
- **But the maximum infection path reaches only 3% to 7% of the population**
  - This is the protective effect of partnership duration and sequence

This level of concurrency doubles epidemic potential

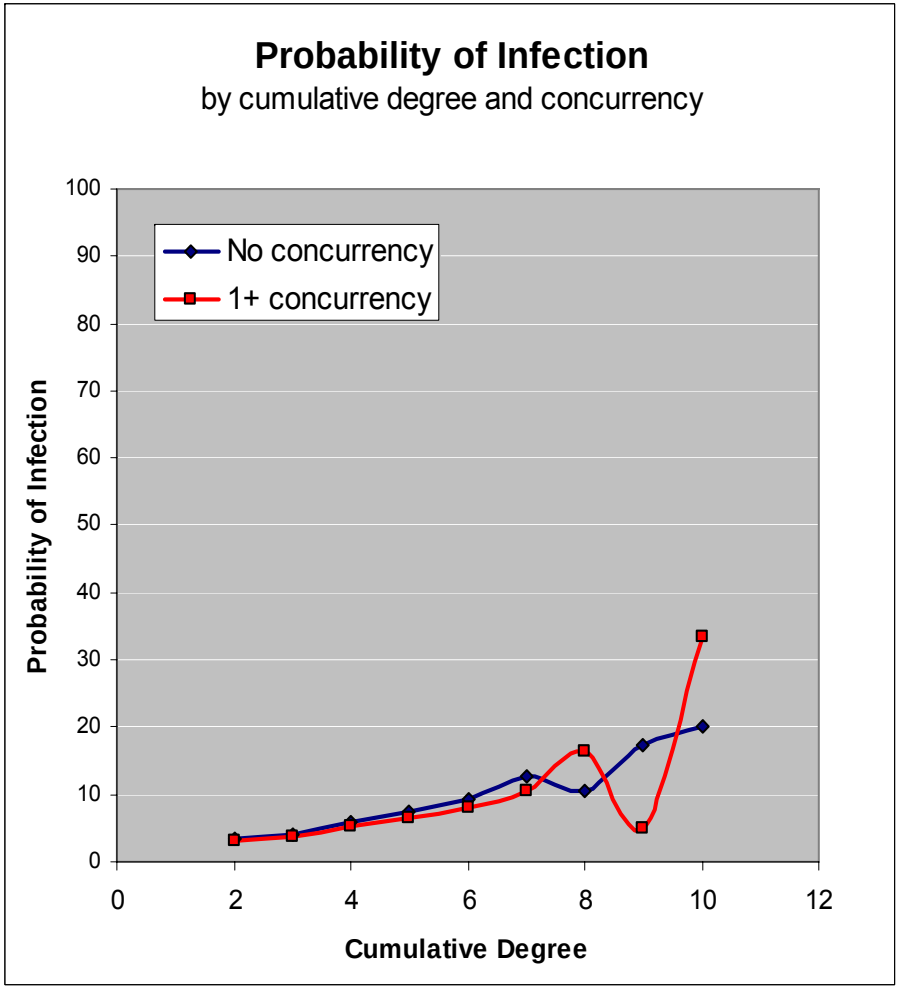


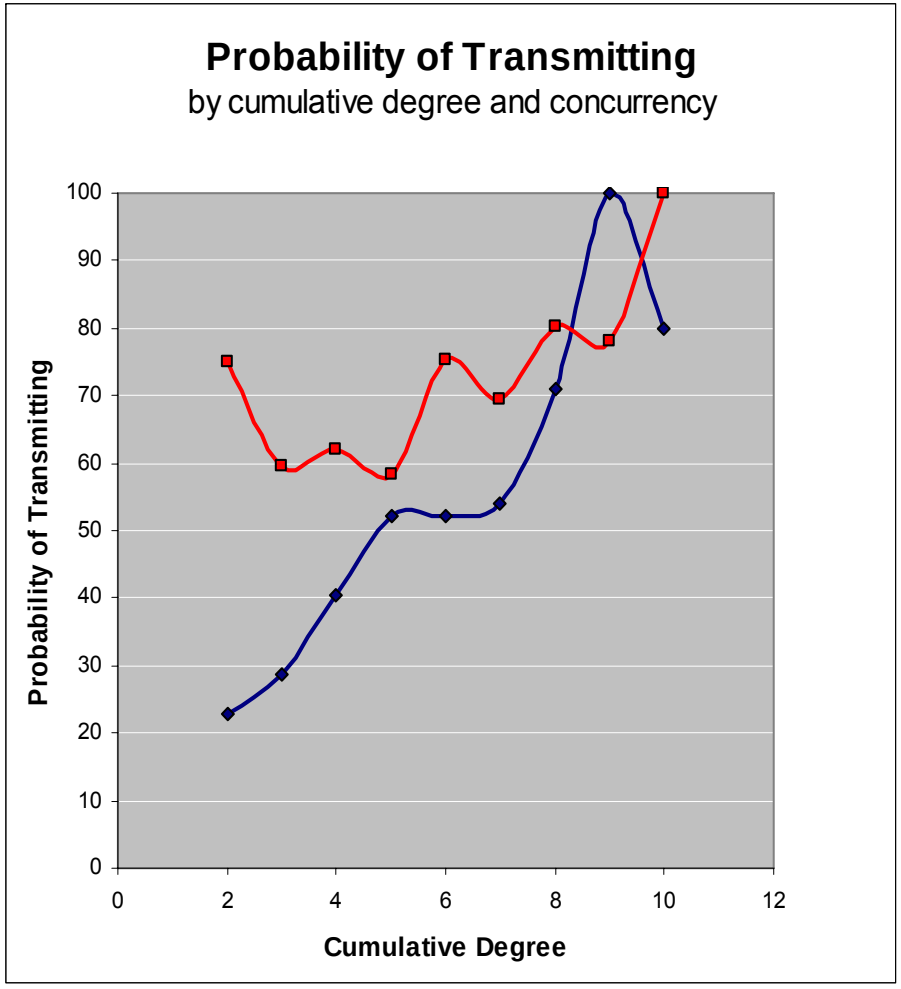
Concurrency is responsible for half of all infections

	Fraction of transmissions through dyads that are
Monogamous	48.5 %
Concurrent	**51.5 %**
Backward chains	19.5
Forward acceleration	30.5
left censored	1.4

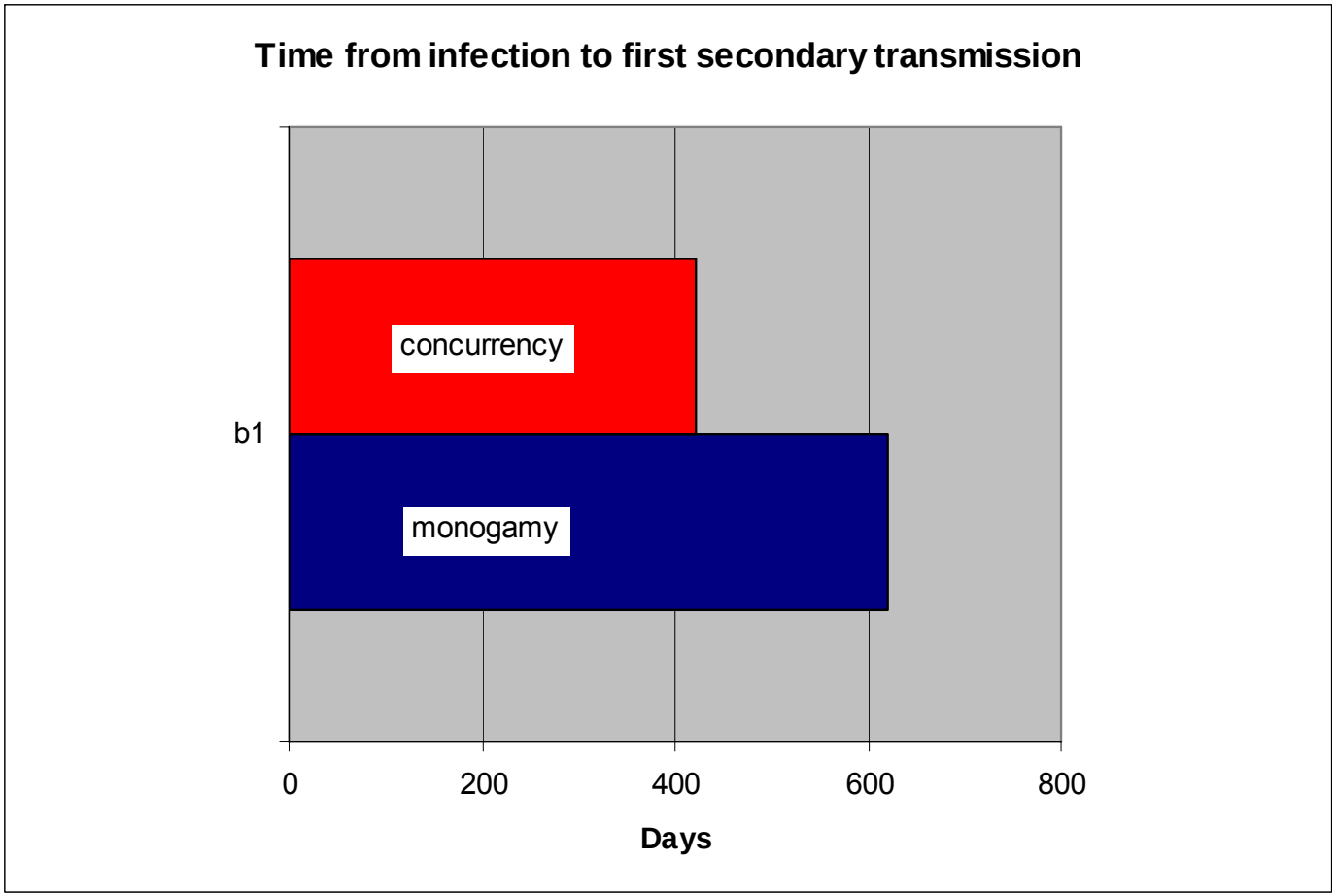
Though only about 5% of dyads are concurrent at any time

At the individual level, concurrency increases the likelihood of transmission, not infection

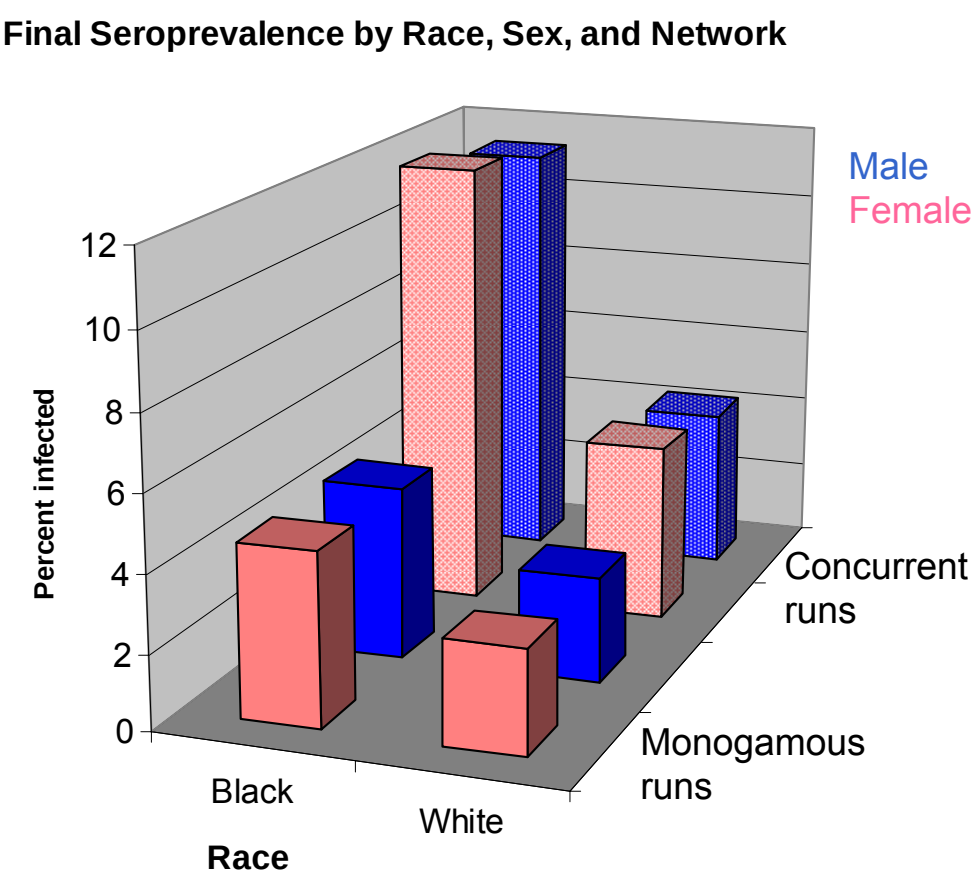




Concurrency reduces the time to transmission by about 1/3

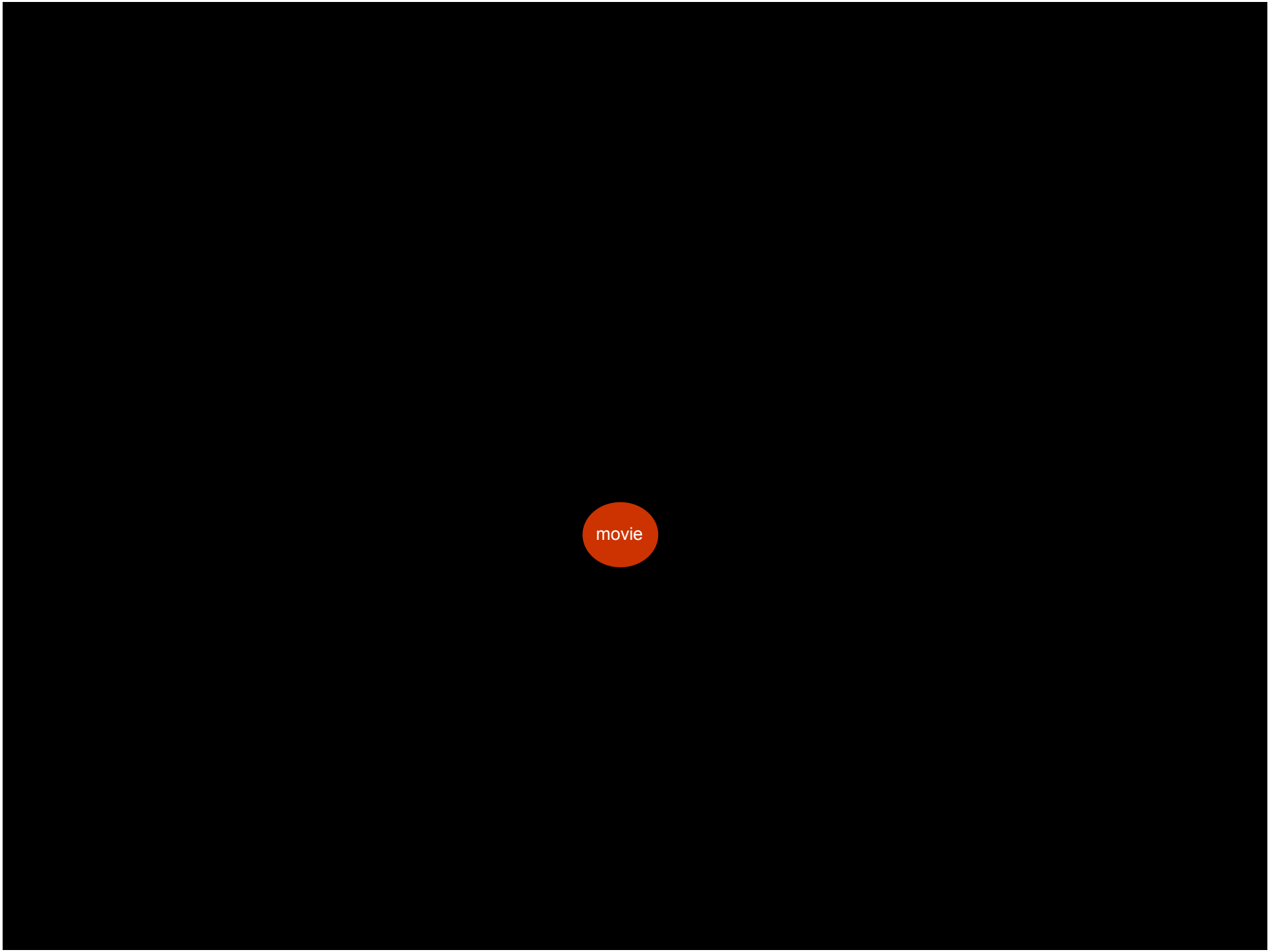


Concurrency increases the racial disparity



Observed levels of concurrency

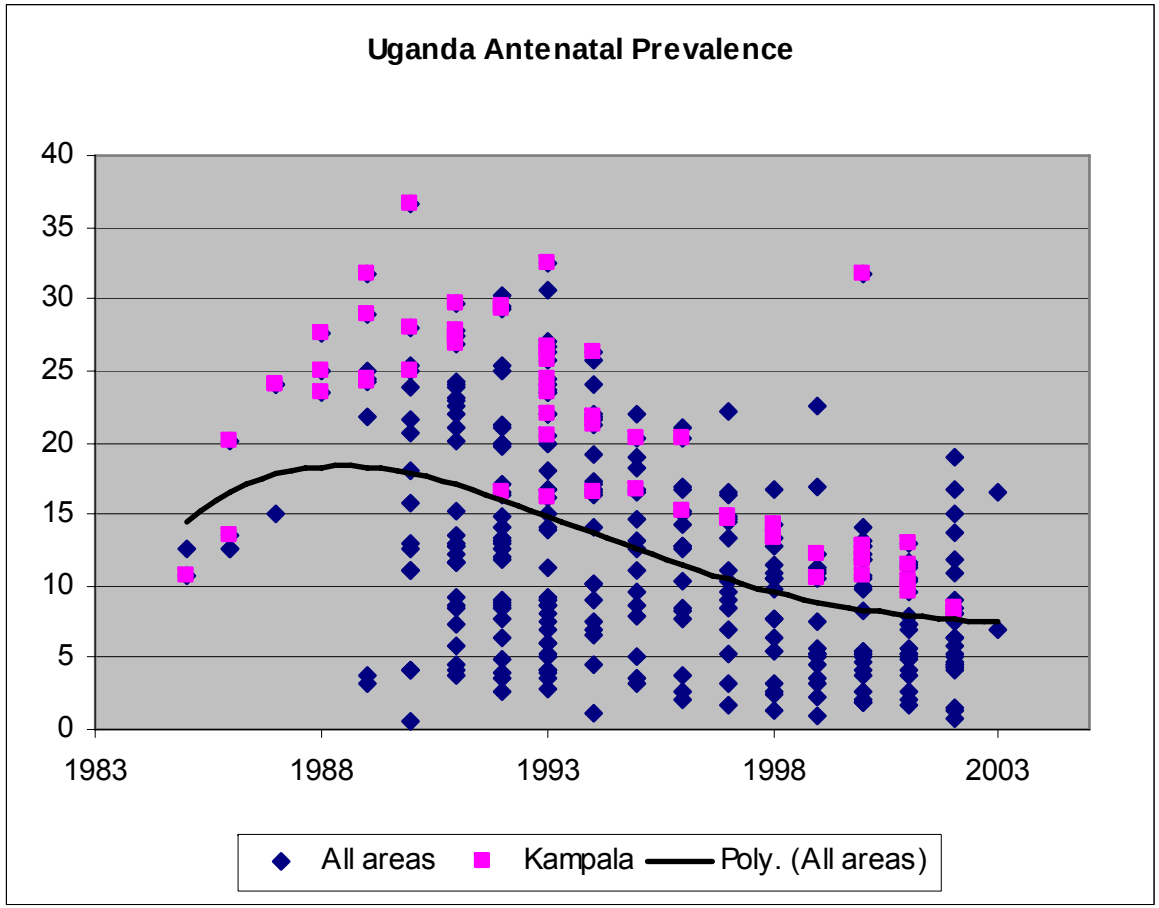
- double the epidemic potential among whites.
- triple the epidemic potential among Blacks


movie

Summary

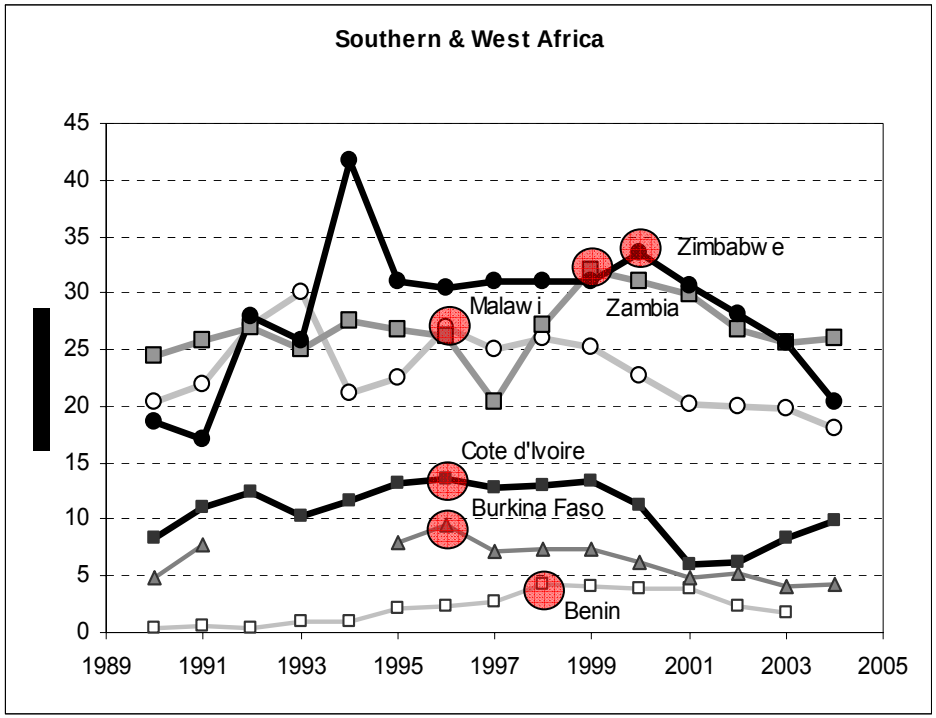
- **The link between individual behavior and population transmission dynamics is mediated by networks**
  - Partnerships play a larger role than generally recognized
  - Per contact transmission probability reduction operate within partnerships
- **Stochastic models provide detailed control of network simulation**
  - Do you need this level of control? It depends on the system.
- **Concurrency may be a major determinant of HIV/STI transmission**
  - Uganda recognized this years ago, and developed an effective prevention message around "*zero grazing*"
  - The "B" in ABC
  - A small change may be enough to bring the HIV epidemic under control

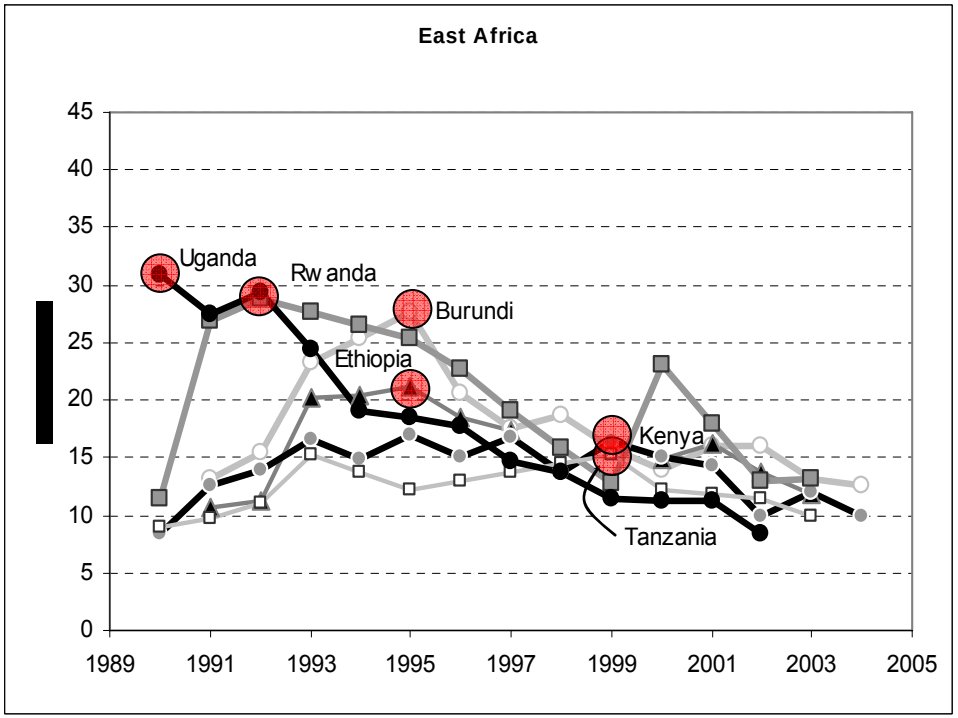
The future of HIV is already happening in Uganda



And it is a hopeful future.

Prevalence declines are in fact happening throughout Sub Saharan Africa





Data from UNAIDS 2006 report : 13 countries have prevalence declines
Median decline is 50%
Predates circumcision, and drugs

Acknowledgments

- **Would like to thank NICHD and NIDA for their commitment to this specific research program over many years.**
  - R29 HD034957
  - R01 HD38210
  - R01 HD41877
  - R01 DA012831

- **The center grants that supported the research environment**
  - Population centers at Penn State and the University of Washington (NICHD)
  - Center for AIDS Research at UW (NIAID)

- **And the data used in this presentation:**
  - The National Longitudinal Survey of Adolescent Health (JR Udry, PI, NICHD)

Readings and resources

Introduction to epidemic modeling (compares deterministic and stochastic approaches)

- Epidemic Modelling: An Introduction. (1999) Daley, DJ and J Gani. Cambridge Studies in Mathematical Biology series #15. Cambridge: Cambridge University Press.

General review of networks and HIV

- "Sexual Networks, Concurrency, and STD/HIV." (2007). Morris, M., S. Goodreau, and J Moody. Sexually Transmitted Diseases. K. K. Holmes, P. F. Sparling, W. E. Stammet al. New York, McGraw-Hill**:** chapter 7.

Tools for network estimation/simulation: **statnet**

- http://statnetproject.org/
- the **statnet** program produced the network movie in the application section of the talk