

Respondent driven sampling for highly at risk populations

Keith Sabin, PhD, MPH
Senior Epidemiologist
Strategic Information and Research Team
HIV Department
World Health Organization



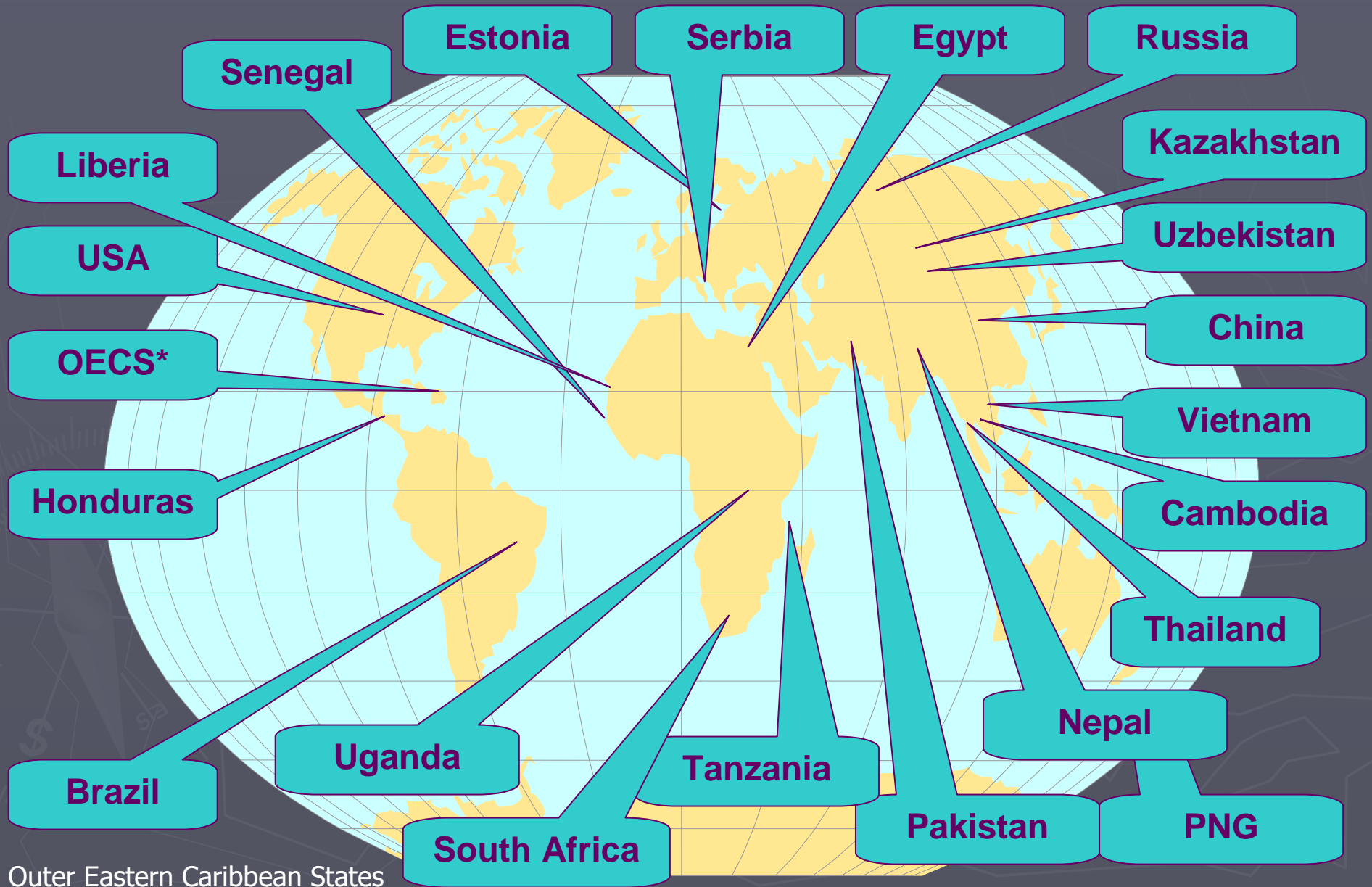
Outline

- I. Why and when to use respondent driven sampling (RDS)?
- II. Implementation
- III. Testing some assumptions
- IV. Analysis
- V. Challenges

Does RDS Work?

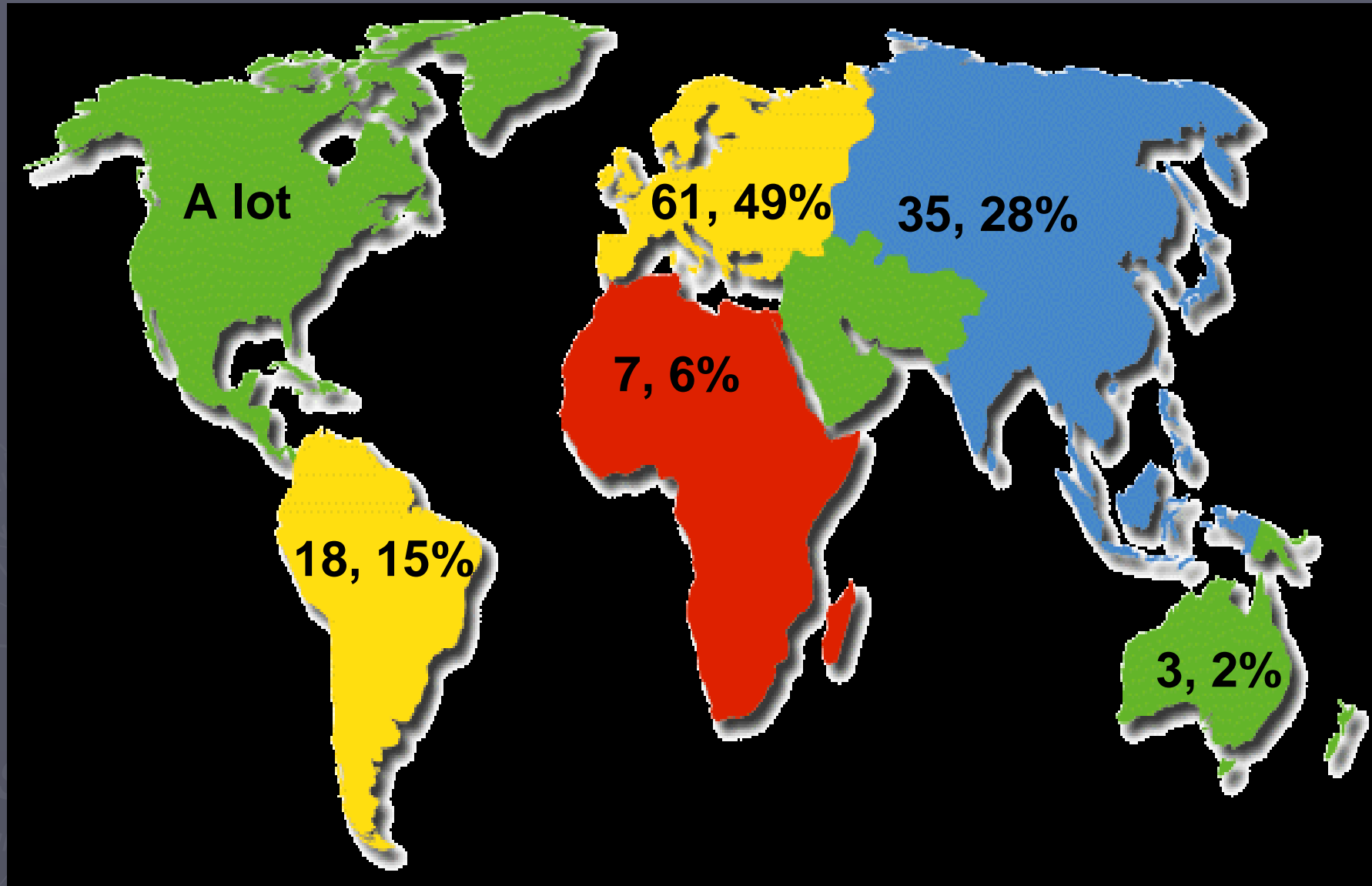


RDS in the World



* Outer Eastern Caribbean States

Where RDS is being Conducted





Honduras
(x4)

Brazil

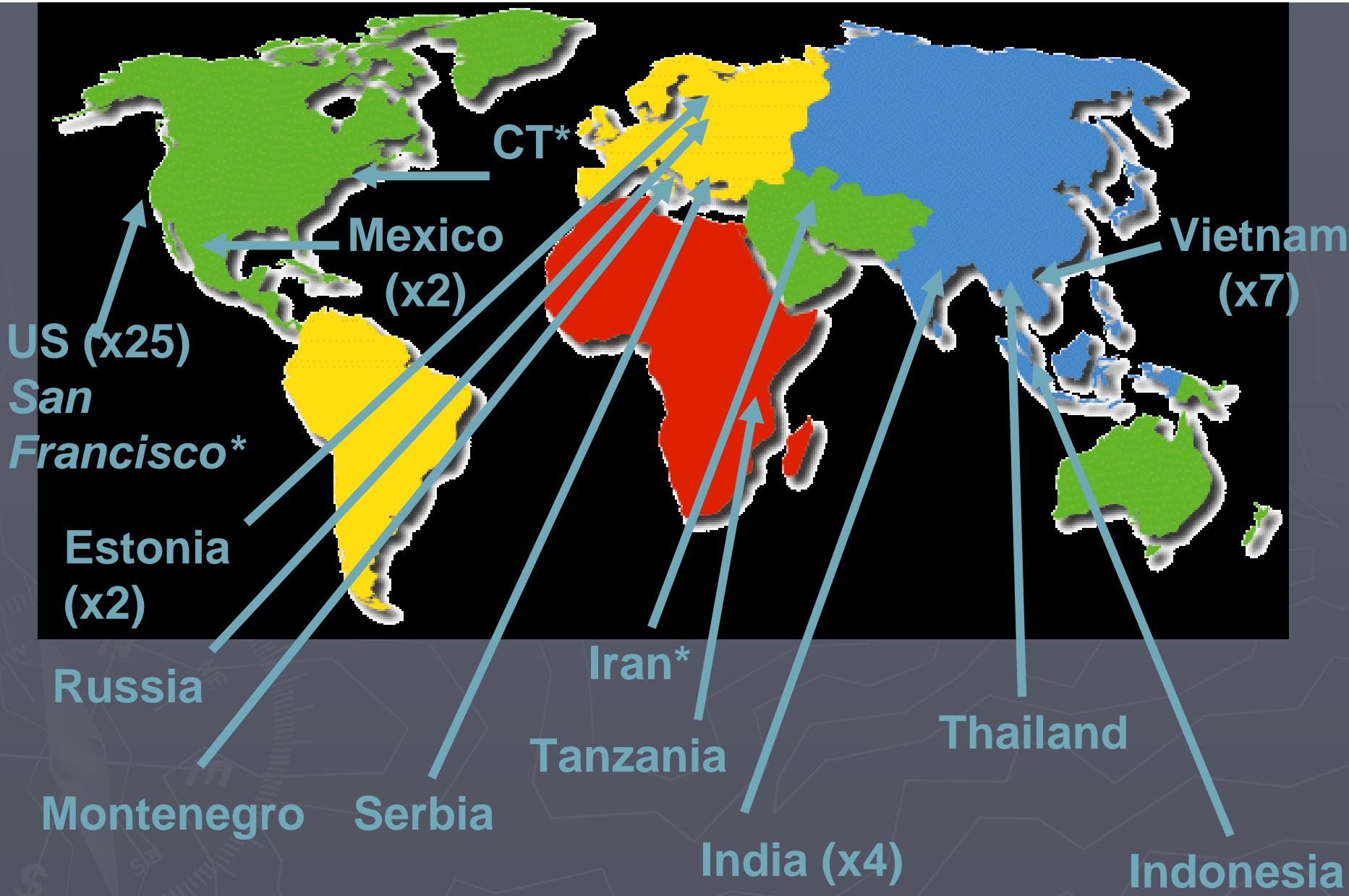
Tanzania

India
(x2)

Vietnam
(x2)

Papua New Guinea
(x2)

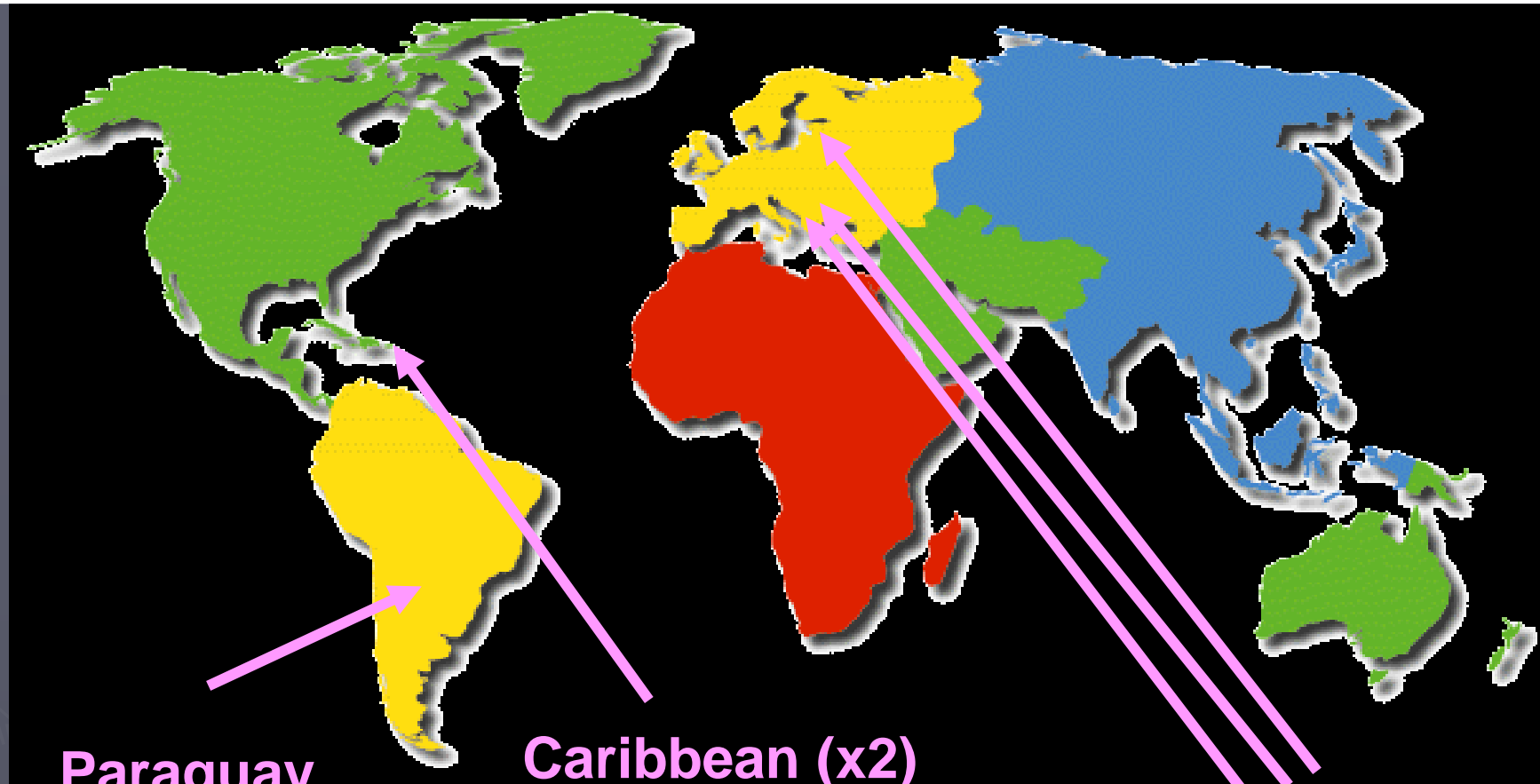
Female sex workers: Successes



IDU: Successes



MSM: Successes



Paraguay

Not enough time
Not well networked
Fear of disclosure

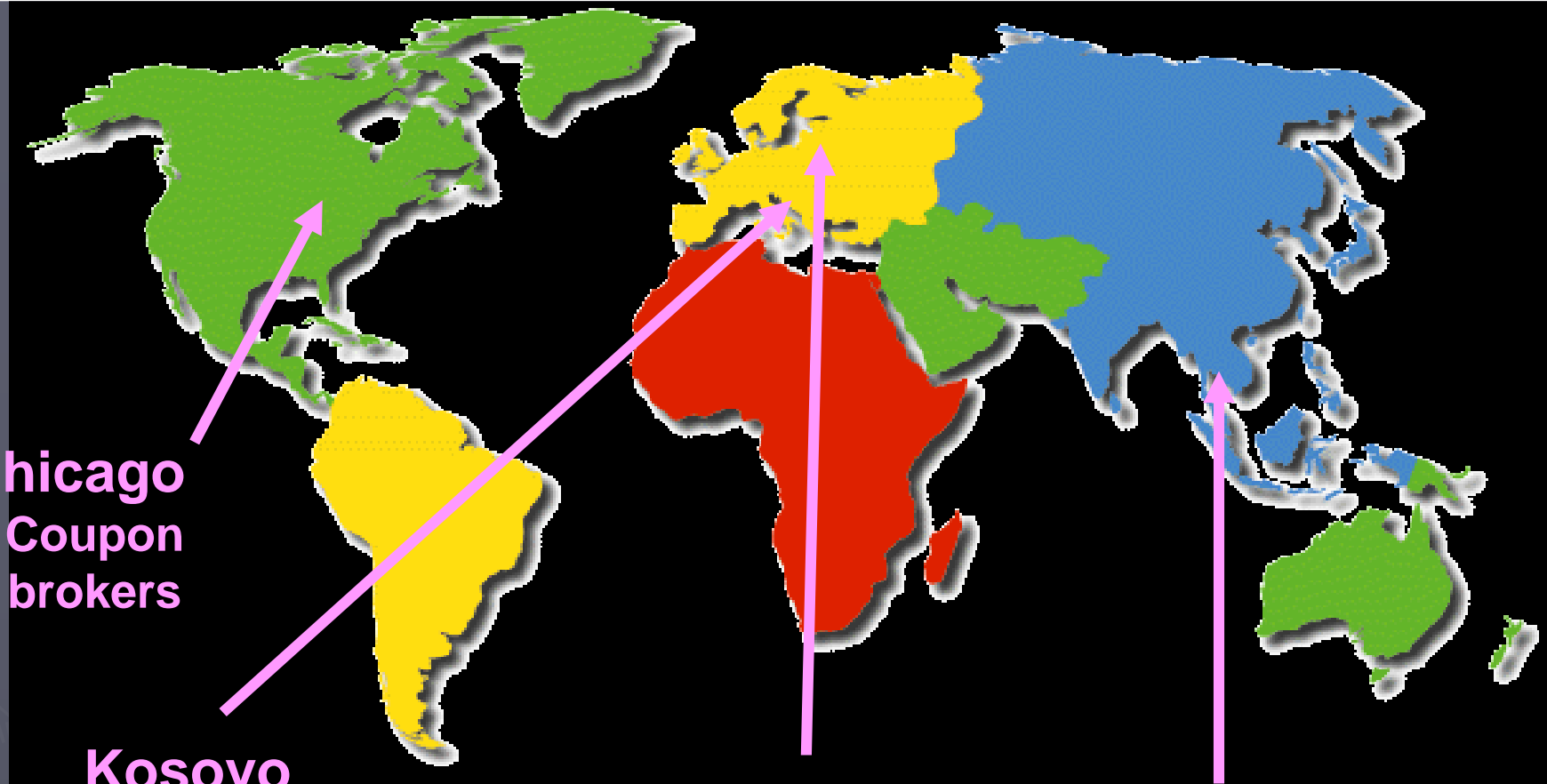
Caribbean (x2)

High stigma
Low incentive

**Serbia,
Montenegro,
Russia**

Too controlled by
pimps

Female sex workers: Failures



Chicago
Coupon
brokers

Kosovo

Russia

Cambodia

Uncertain if
really RDS?

Uncertain if
really RDS?

Coupon
brokers

Network size?

No adjustment

Linkage?

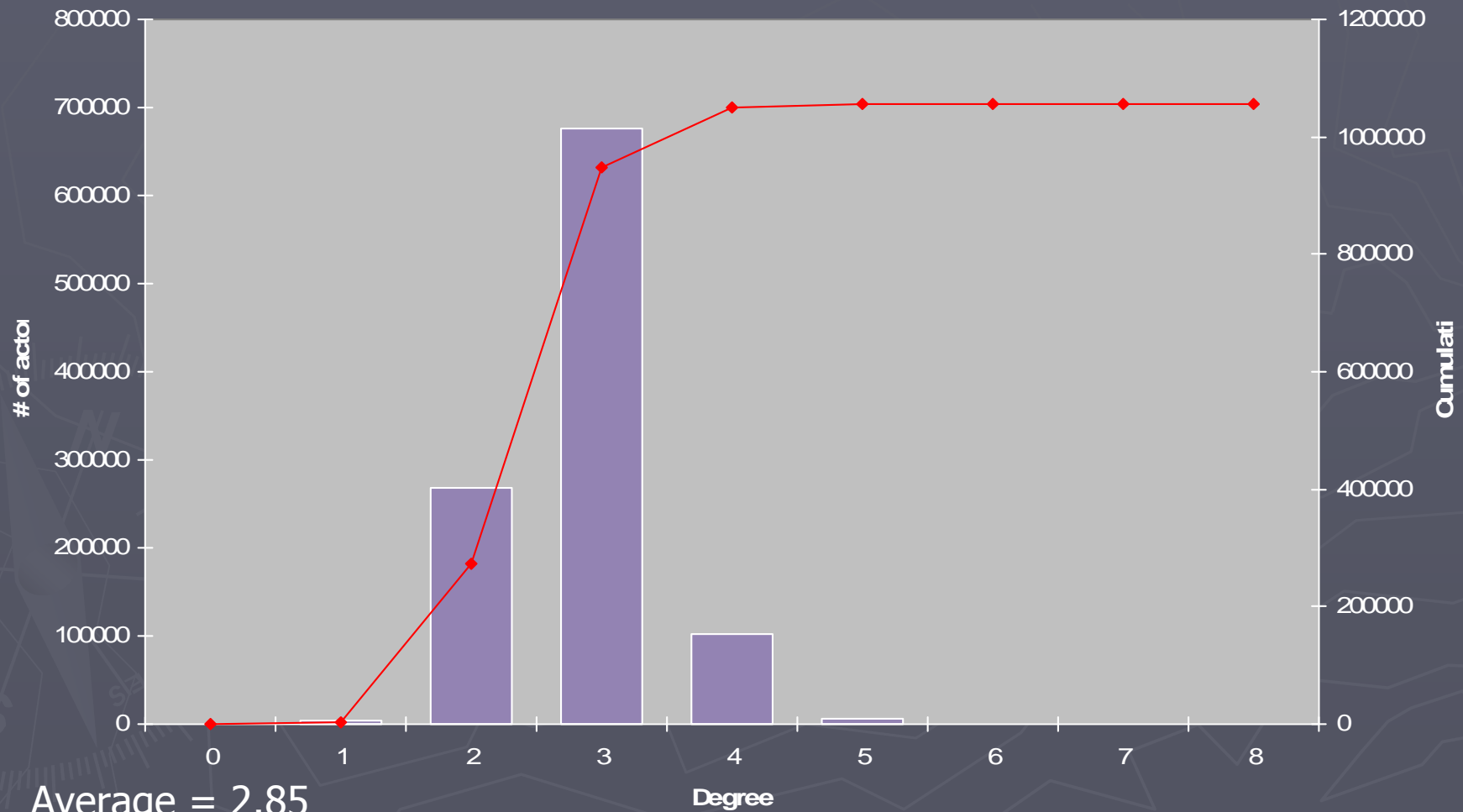
IDU: Failures



Why does RDS work?



Distribution of Actors Appearing in Films with Gerard Depardieu



Average = 2.85

Total # linkable = 3,012,634

Depardieu Numbers

▶ Kevin Bacon (2)

- *Frost/Nixon* > Clint Howard in *Unhook the Stars*

▶ Jerry Lewis (2)

- *Funny Bones* > Richard Griffiths in *Vatel*

▶ Carla Bruni (2)

- *Paparazzi* > Didier Cauchy in *Asterix et Obelix*

▶ Charlie Chaplin (2)

- *Countess from Hong Kong* > Geraldine Chaplin in *I want to go home*

Advantages

- ▶ Yields probability sample, theoretically
 - Statistically valid estimates with variance
- ▶ Reaches populations which are difficult to access
- ▶ Maintains safety of staff
- ▶ Offers opportunity for interventions
- ▶ Cost effective

Limitations

- ▶ Population must be a network
- ▶ Must be able to verify group membership
- ▶ Must track links between recruiters and recruits-coupon management
- ▶ Incentives (usually)
- ▶ Very difficult to deal with selective non response bias
- ▶ Analysis not fully worked out, difficult

Theoretical Assumptions

- 1) Respondents know one another as members of the target population
- 2) Respondents' network forms a single large component
- 3) Sampling occurs with replacement
- 4) Respondents can accurately report their personal network size.
- 5) Respondents recruit as though they are sampling randomly from their personal networks

Functional Assumptions

1. Bias from seeds is eliminated in a few waves
2. Equilibrium is approximated in a few waves
3. Wave of recruitment is unimportant - unit of analysis is recruiter-recruit dyad
4. Recruitment chain of a respondent is unimportant - unit of analysis is recruiter-recruit dyad
5. Permitting multiple recruitments does not introduce bias or increase variance of point estimates

Implementation



Logistics

- ▶ Formative work
- ▶ Safety (controllable environment)
- ▶ ACASI
- ▶ Specimen collection
- ▶ Overlapping staff
- ▶ Service provision
- ▶ Requires vigilance against fraud (if incentives used)





PROJETO AMAA

Ministério
da Saúde 

Endereço: Rua Taiqueirinha, Nº 57-Centro,
(próximo ao ...)

Fone:
(você pode telefonar para agendar a entrevista)

1

Este cupom é válido até:

DD/MM/AAAA

ID:

Horário de atendimento:

13:00 as 18:00h (de segunda a sexta)

ESTE CUPOM DEVERÁ SER ENTREGUE NO DIA DA ENTREVISTA

PROJETO AMAA

Endereço: Rua Taiqueirinha, Nº 57-Centro,
(próximo ao ...)

Fone:
(você pode telefonar para agendar a entrevista)

2

Este cupom é válido até:

DD/MM/AAAA

ID:

Horário de atendimento:

13:00 as 18:00h (de segunda a sexta)

CONVITE

Você está sendo convidada a participar do Projeto: AMAA.

Se você for "Batalhadora" na cidade de Manaus e tiver mais de 14 anos de idade, venha nos visitar e traga este cupom.

Ele não pode estar rasgado, rasurado ou sem o número de identificação e a validade vencida.

Venha nos conhecer, sua presença é importante!



PROJETO AMAA

Traga este cupom após uma semana da data de vencimento.

Ele não pode estar rasgado, rasurado ou sem o número de identificação.











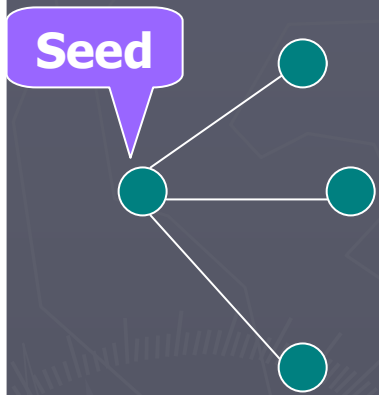
Wave 1

Wave 2

Wave 3

Wave 4

Wave 5



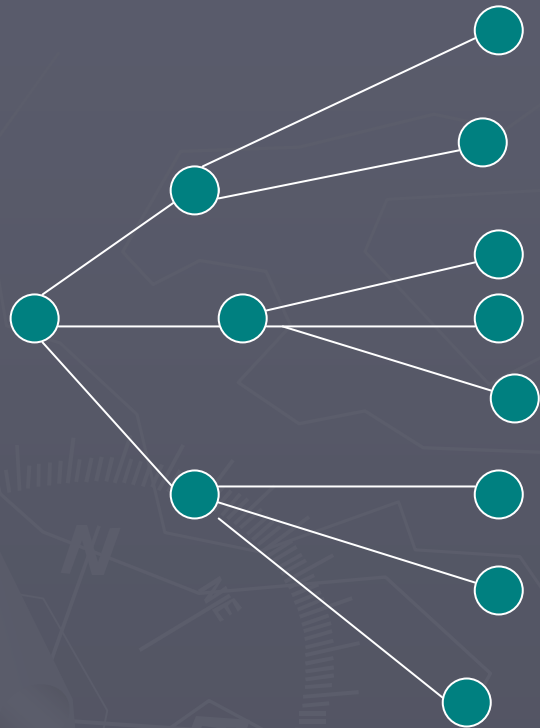
Wave 1

Wave 2

Wave 3

Wave 4

Wave 5



A Probability Method? (I)

Assumptions:

1. Respondents recruit people with whom they have a relationship.
2. Recruits are randomly sampled from personal networks
 1. Assumes meets criteria for Markov Chain
3. Social network sizes are “know-able”

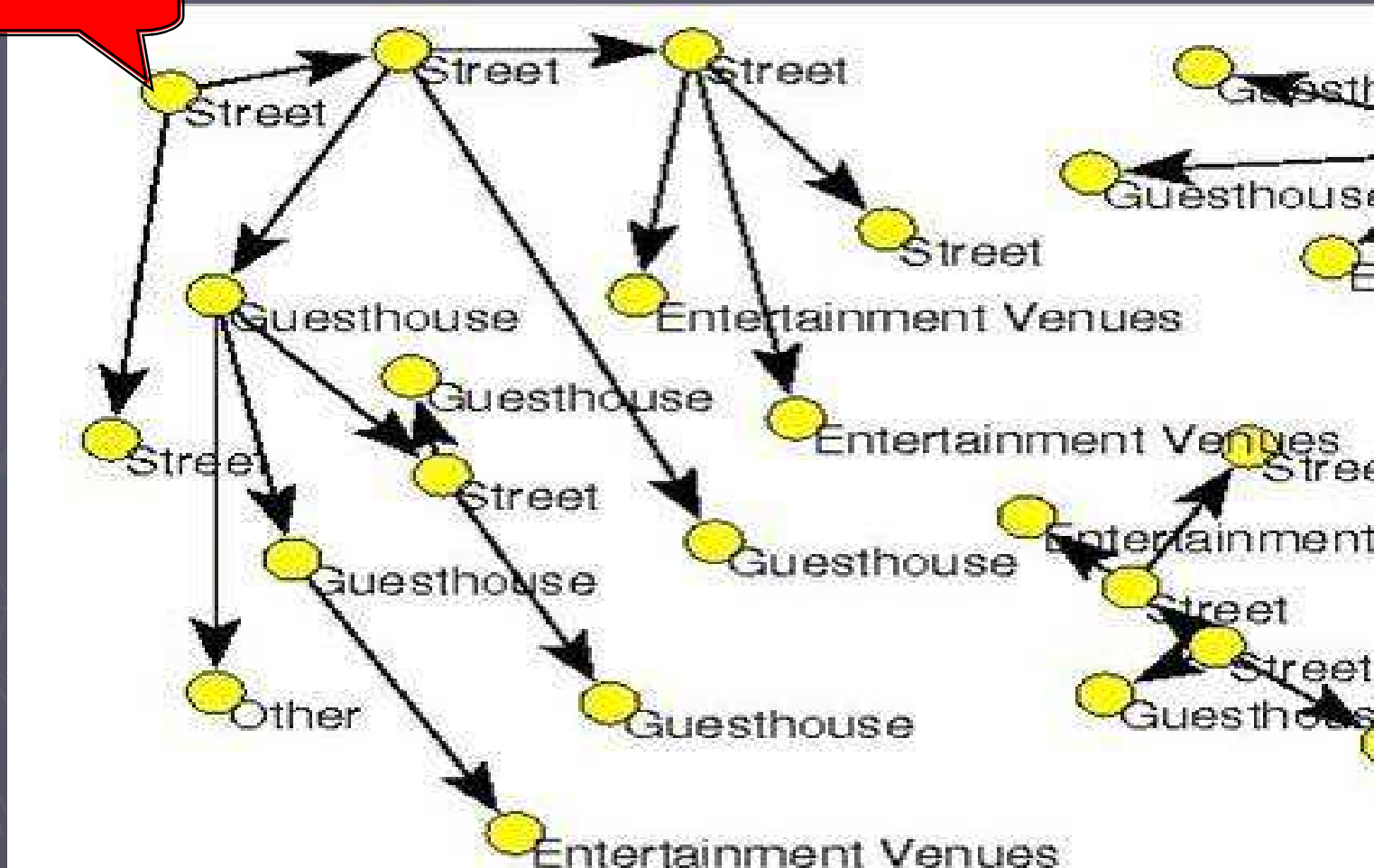
A Probability Method? (II)

Assumptions:

1. Respondents recruit people with whom they have a relationship. **Yes, mostly**
2. Recruits are randomly sampled from personal networks **Yes, sort of**
 1. Assumes meets criteria for Markov Chain
3. Social network sizes are “know-able”
Probably not very well

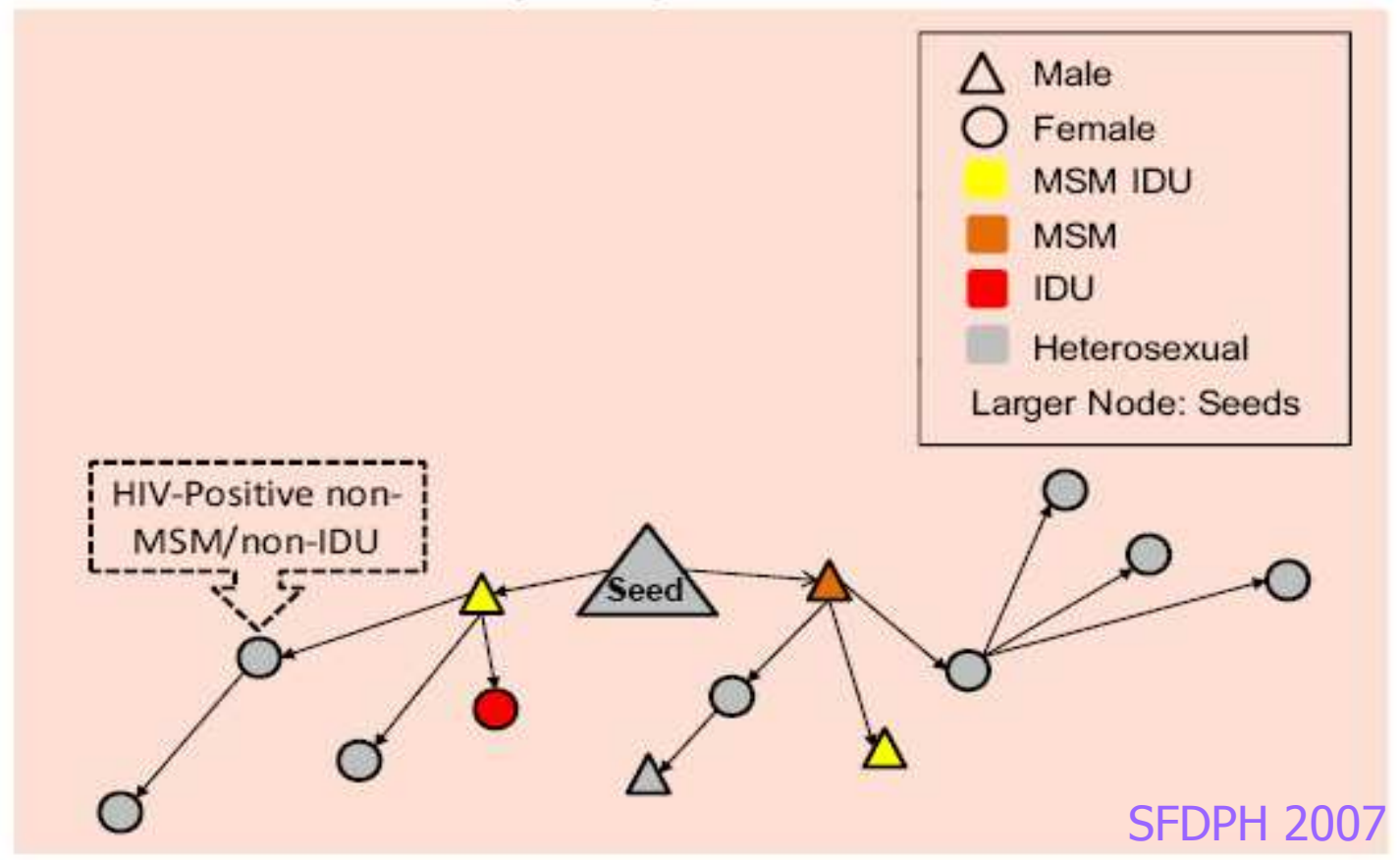
Do seeds matter?

Seed



Unquantifiable Advantages of RDS

Figure 10.3 Recruitment Chain, Single Heterosexual HIV Positive, 2006, San Francisco



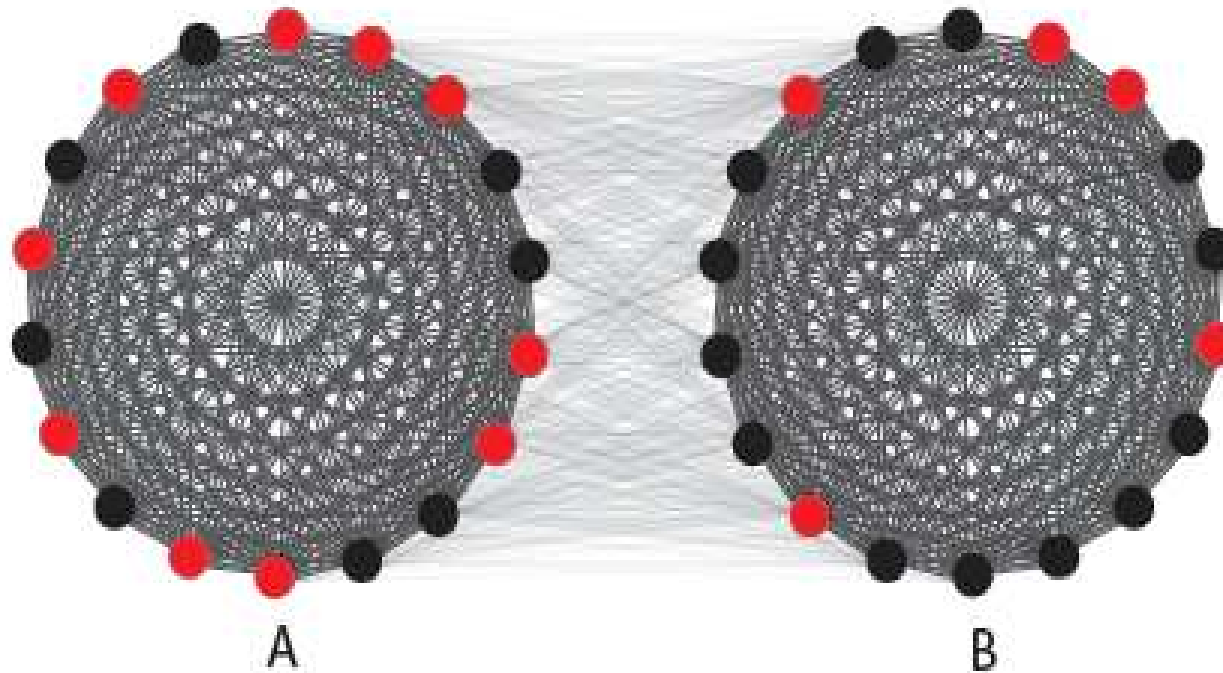
Analysis

- ▶ Specialized software (RDSAT)
 - Steep but short learning curve
 - Does not save code (time consuming)
 - Weights exportable
- ▶ Asymptotically unbiased
- ▶ Adjusts for homophily and network size
- ▶ Yields variance for confidence intervals
- ▶ R can also be used

Analysis: Weighting

- ▶ Horvitz-Thompson estimator
- ▶ Dual component weight adjusts for
 - Homophily
 - Degree size

Homophily



Two distinct networks with few connections
(Red dots depict infection)

Goel and Salganik 2007

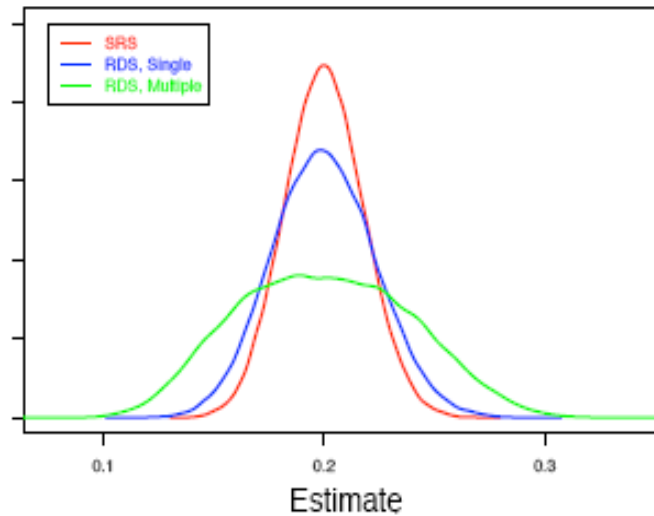
Challenges



Unresolved Issues

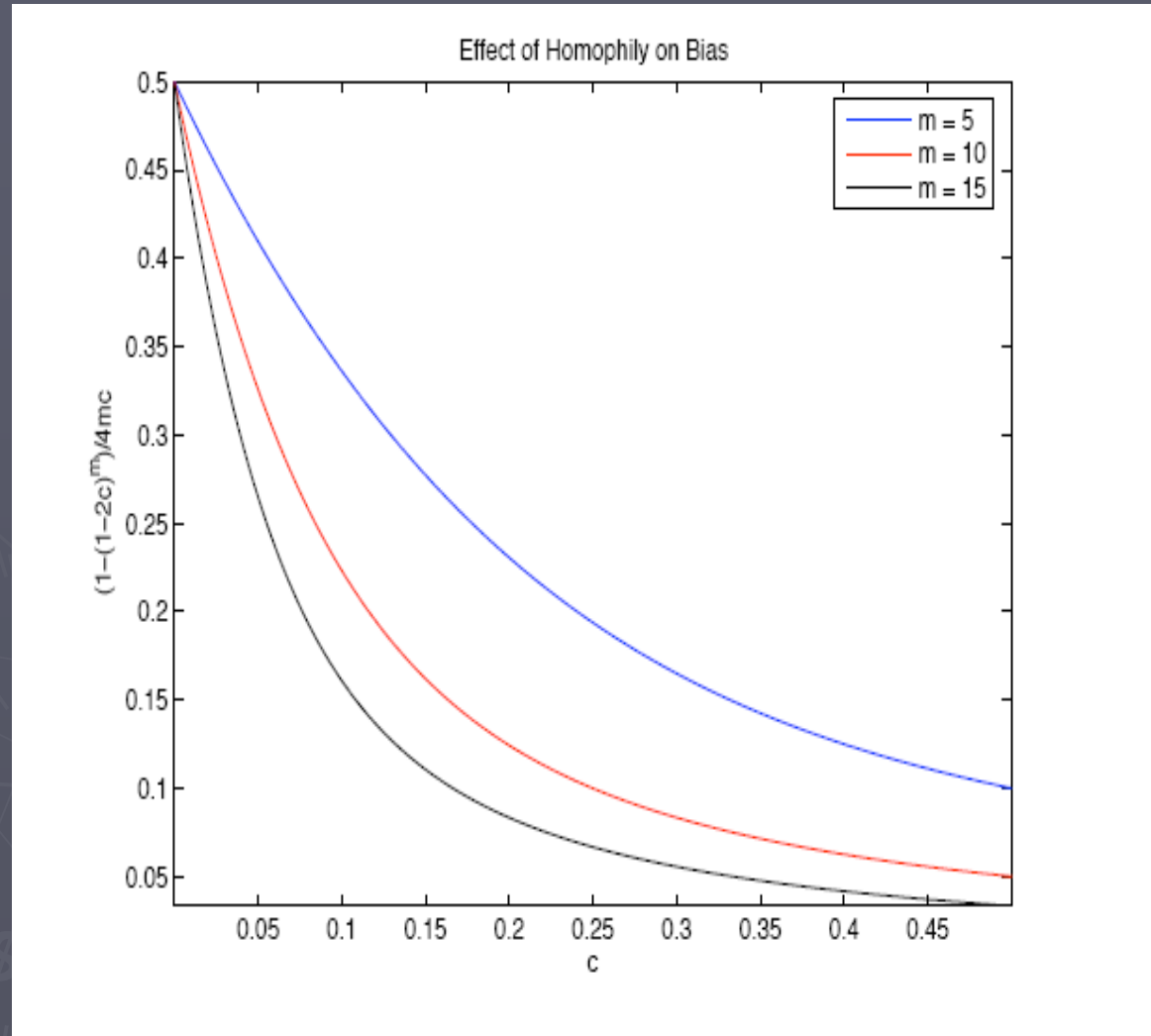
- ▶ Design effect may be high
- ▶ Assumptions may be very important
 - Often violated
 - So, how bad is bad when violated?
- ▶ Can a single weight be generated for all analyses?
- ▶ Valid ascertainment of peer network size (degree)

Comparing sampling schemes (n=500)



	Design effect	effective sample size
Simple Random Sampling	1	500
RDS – Single Recruitment	1.5	335
RDS – Multiple Recruitment	3.7	136

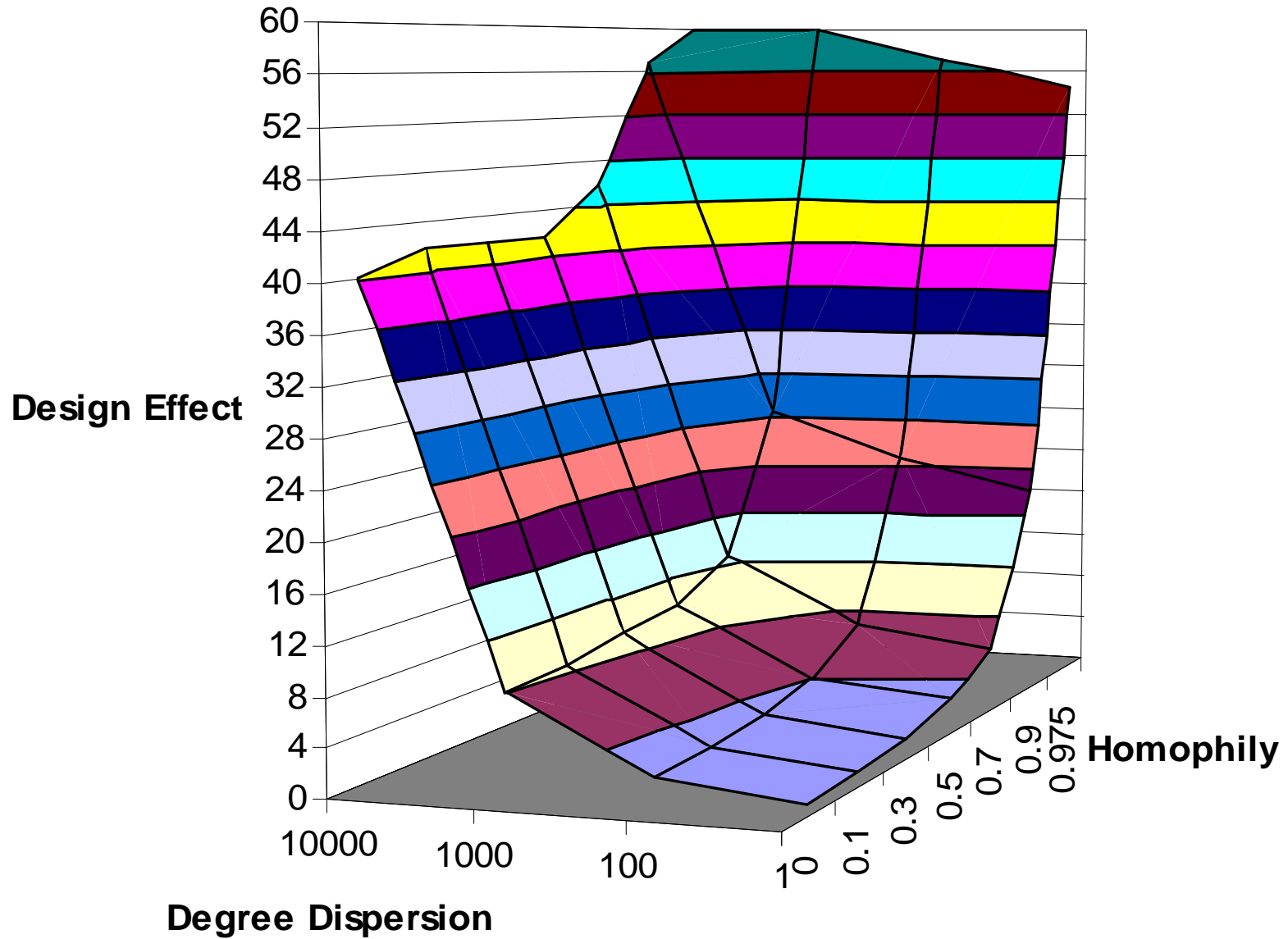
Effect of Homophily



**Design effect widens
CI by 1.5x.**

**Therefore, a sample of
500 -> 355.**

Design Effect Space



Conclusions

RDS is:

- ▶ an effective recruitment tool
 - For networked populations
- ▶ possibly a probability sampling method
- ▶ not fully formed