

# Lecture 5

## Measuring health in hard-to reach populations



Social Studies of Health 27/11/20



## Sampling and recruitment

- ◉ Sampling and recruitment is always an issue
- ◉ Hard-to reach populations
  - No sampling frame (not institutionalized)
  - Not willing to disclose the group membership (stigma + threats)
  - Might be of special interests for the health research due to the health risks
  - Ignoring leads to no action



## **Why traditional things wouldn't be helpful?**

---

- Subpart of representative sample?
  - Telephone interviews (CATI) to find people who inject drugs/illegal migrants etc?



## Why traditional things wouldn't be helpful?

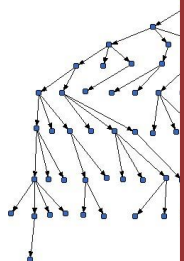
---

- ◉ Subpart of representative sample?
  - Telephone interviews (CATI) to find people who inject drugs/illegal migrants etc?
  - More non-response
  - Not covered by “representative sample” – no phone, other places to live, rarely at home
  - Sensitive question (SDB/intrusiveness/threat of disclosure)



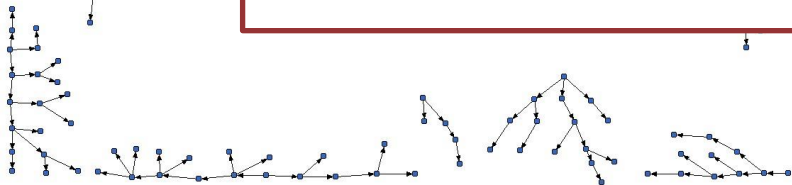
## Two approaches

### Chain-referral

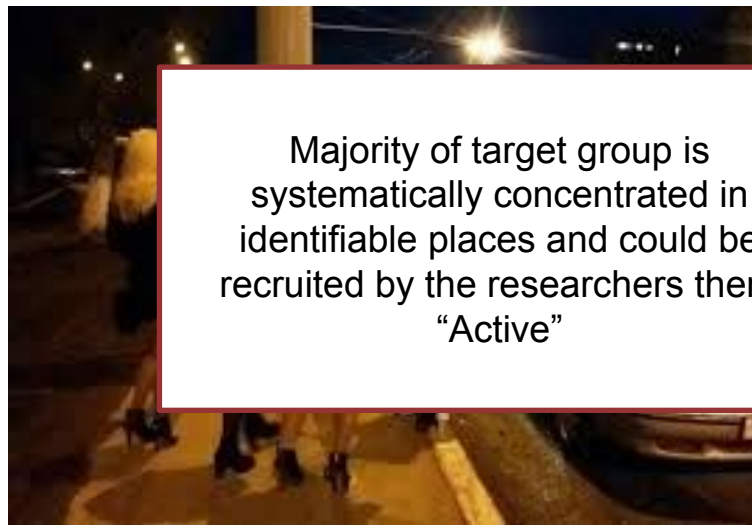


Members of the target group are well-connected and are willing to recruit each other

“Passive”



### Location



Majority of target group is systematically concentrated in identifiable places and could be recruited by the researchers there

“Active”



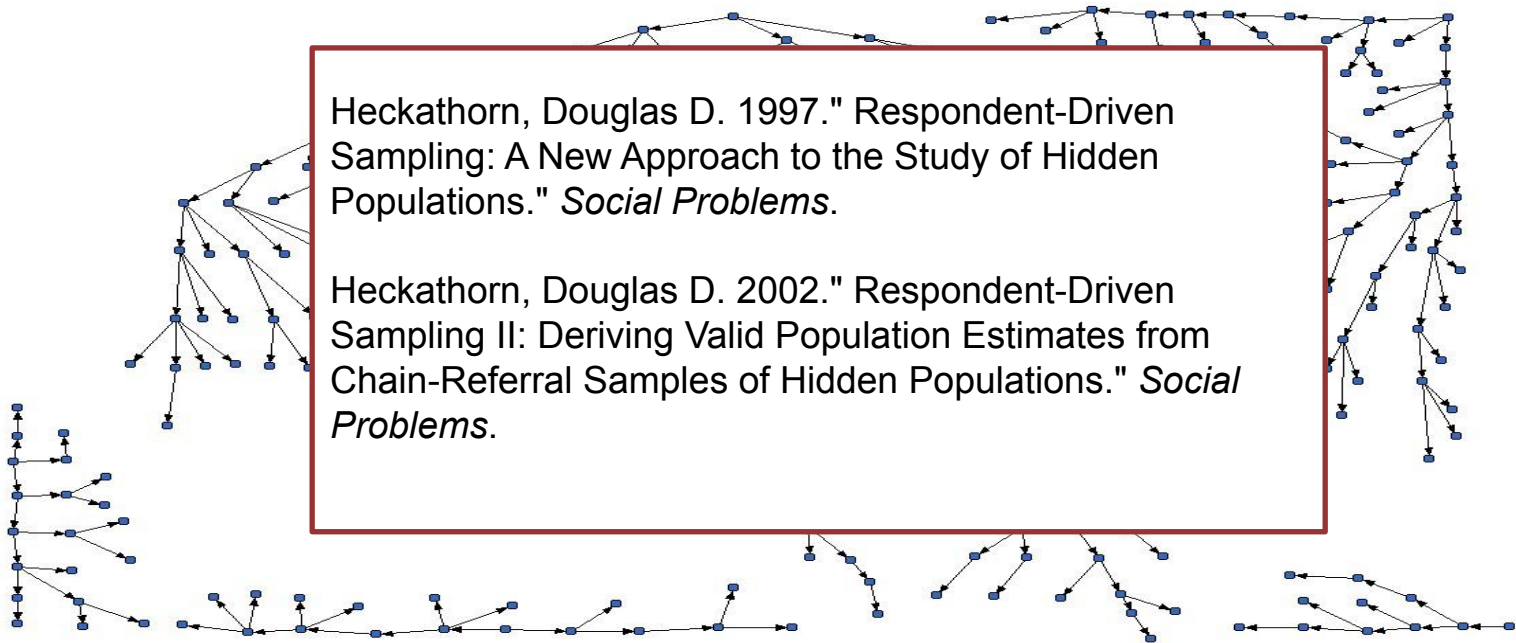
## Chain-referral

---

- ⦿ Snow-ball sampling
- ⦿ Chain-referral sampling
- ⦿ Respondent-driven sampling



## Respondent-driven sampling



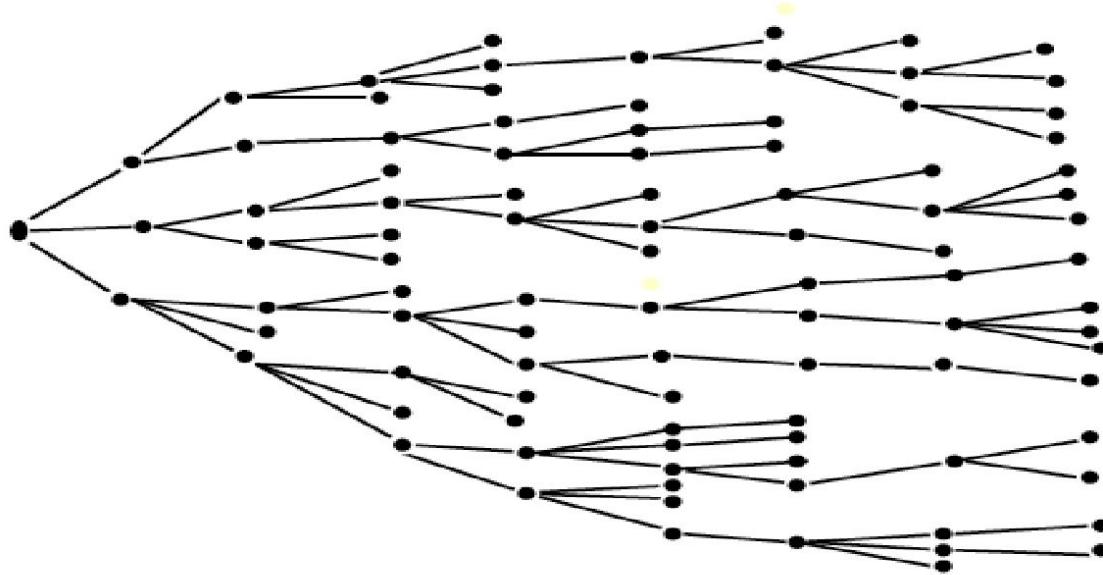
Heckathorn, Douglas D. 1997." Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations." *Social Problems*.

Heckathorn, Douglas D. 2002." Respondent-Driven Sampling II: Deriving Valid Population Estimates from Chain-Referral Samples of Hidden Populations." *Social Problems*.



# Recruitment process and associated terms

Seed Wave1 Wave2 Wave3 Wave4 Wave5 Wave6 Wave7 Wave8

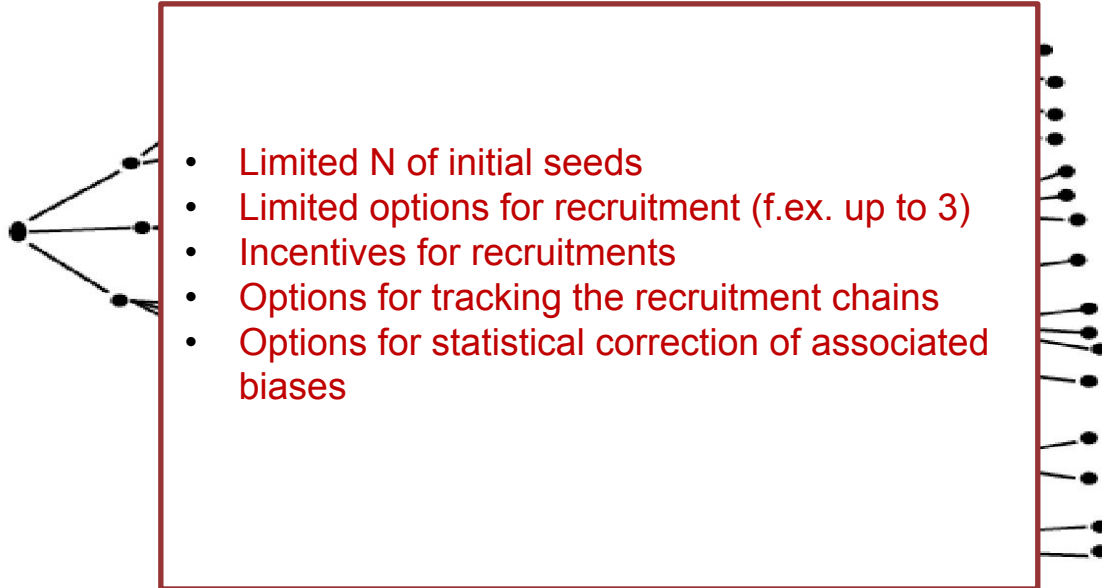






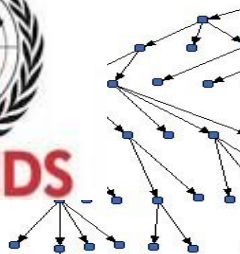
## Recruitment process and associated terms

Seed Wave1 Wave2 Wave3 Wave4 Wave5 Wave6 Wave7 Wave8





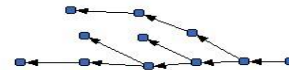
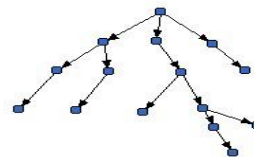
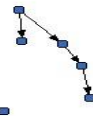
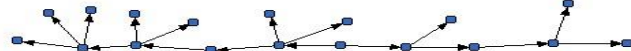
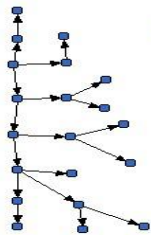
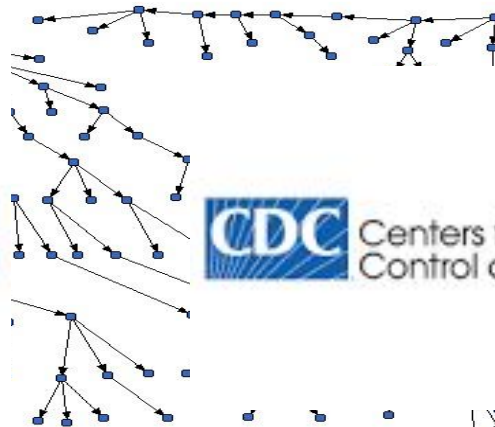
# Respondent-driven sampling



World Health Organization



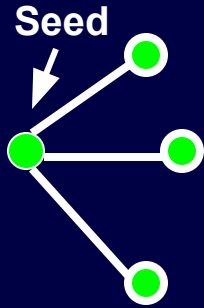
Centers for Disease Control and Prevention



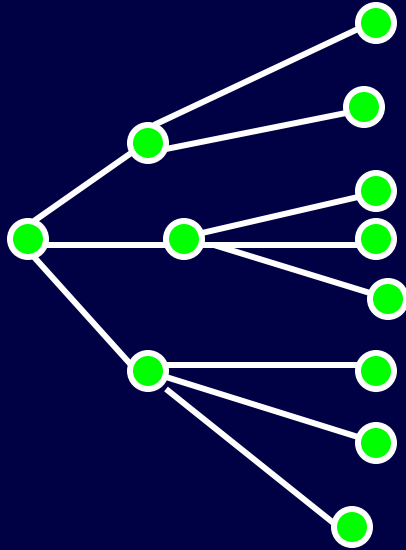
## RDS Methods

- Type of chain referral sampling to reach hidden populations
- Begin with a set of non randomly selected seeds
- Seeds recruit peers, who recruit peers, etc.
- Recruits are linked by coupons with unique identifying numbers
- Recruitment quota through coupons
- Incentives provided for completed survey and for each successful recruit

Wave 1   Wave 2   Wave 3   Wave 4   Wave 5

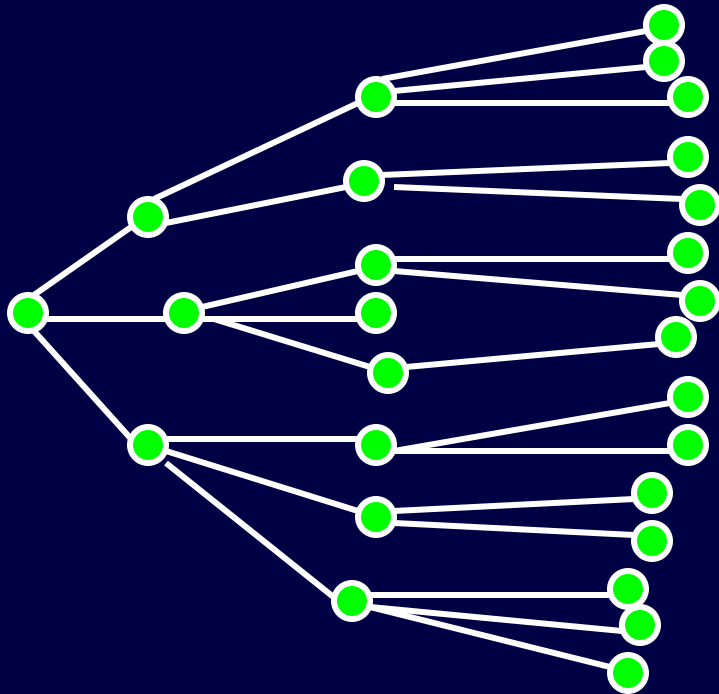


Wave 1   Wave 2   Wave 3   Wave 4   Wave 5



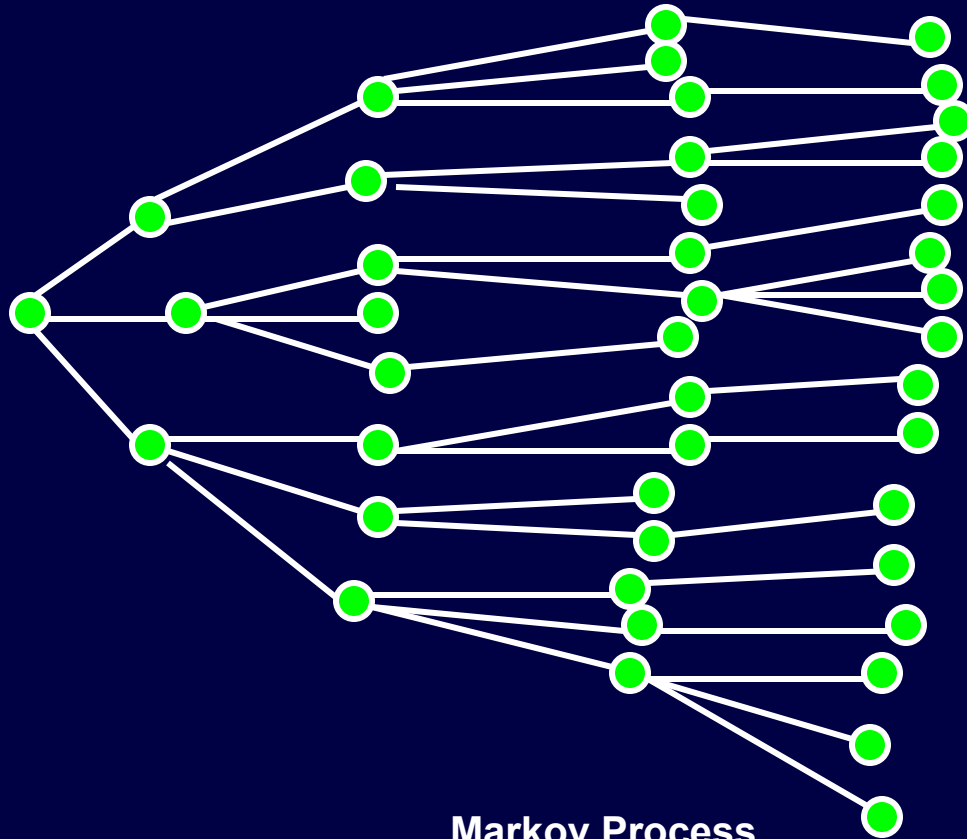
Markov Process

Wave 1   Wave 2   Wave 3   Wave 4   Wave 5



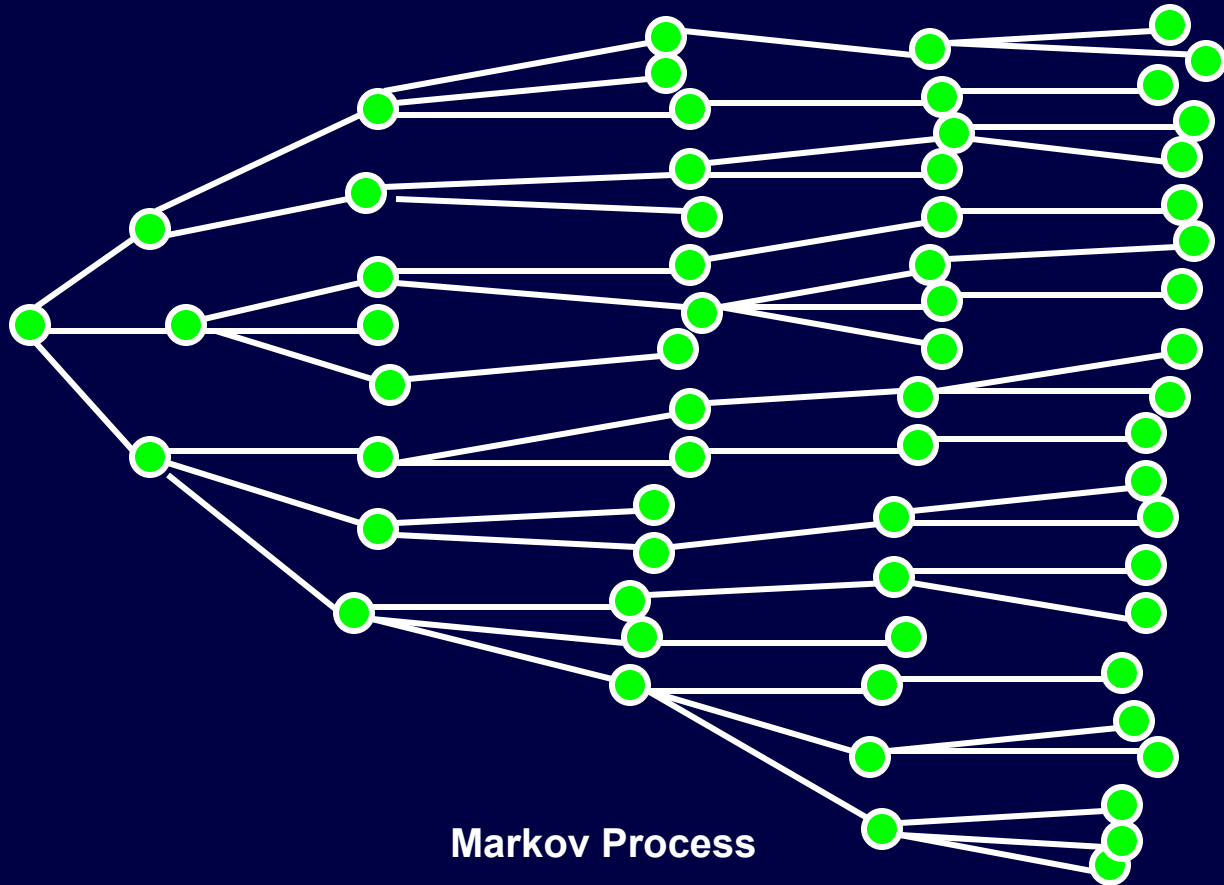
Markov Process

Wave 1   Wave 2   Wave 3   Wave 4   Wave 5



Markov Process

Wave 1 Wave 2 Wave 3 Wave 4 Wave 5





## Important Terms-RDS Methods

Seeds

Wave

Chain

Primary incentive

Secondary incentive

## **Steps involved in RDS**

- ◉ Begin with a set of non randomly selected seeds
- ◉ Seeds recruit peers, who recruit peers, etc.
- ◉ Recruits are linked by coupons with unique identifying numbers
- ◉ Incentives provided for completed survey and for each successful recruit

## The Theory Behind RDS

- Long referral chains
- Final sample will be independent of those selected as “seeds”
- Final sample will be similar to the population of the network from which you are recruiting

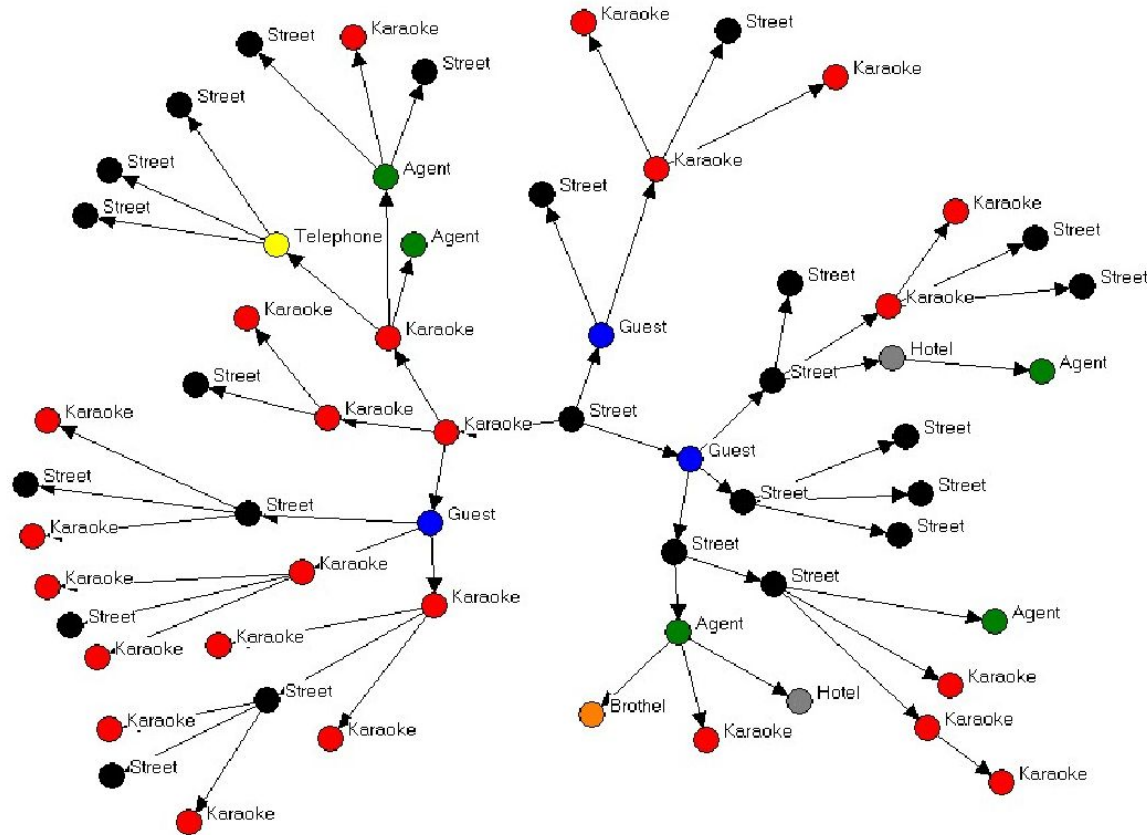
## **You cannot do RDS If:**

- The members of your target population **ARE NOT** well networked (need some formative research)
- The members of your target population are **TOO** stigmatized and afraid to go to your RDS interview site

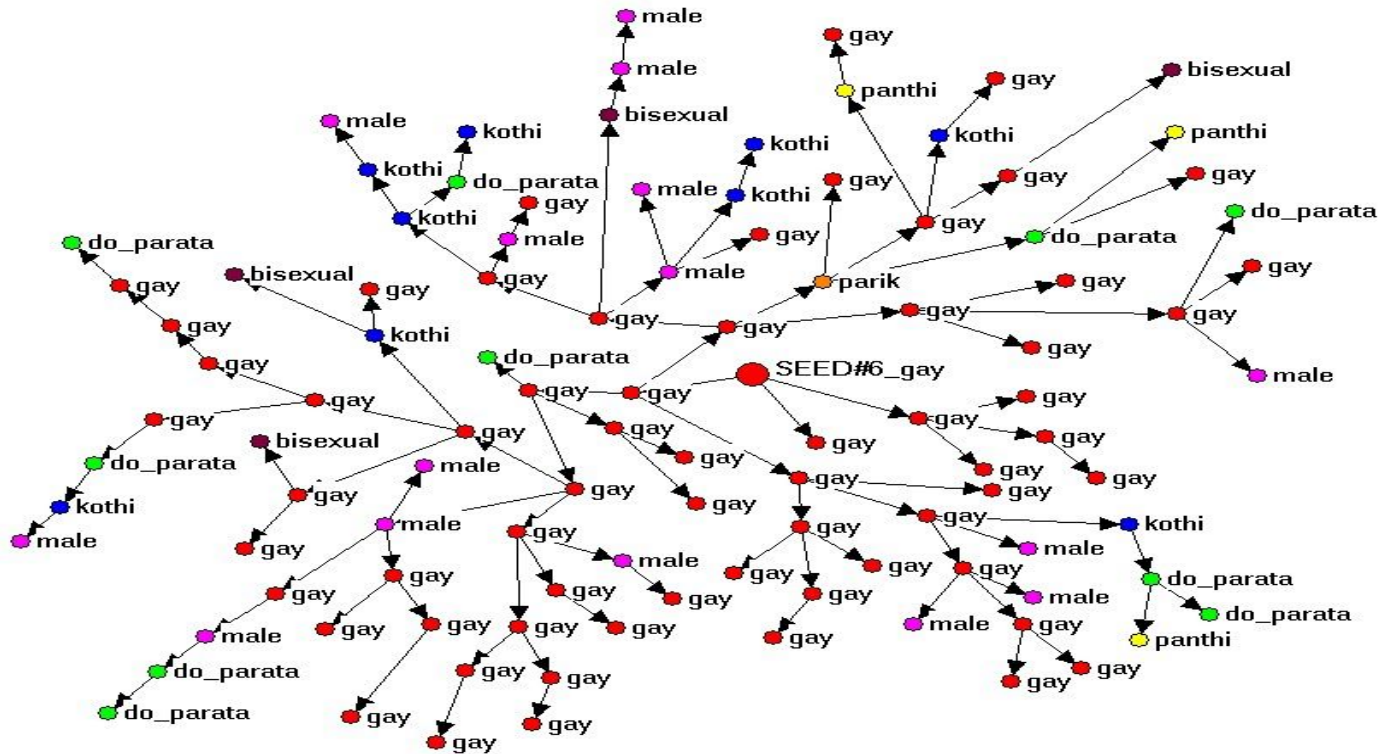
## **Information that MUST be gathered during RDS**

- Personal Network Size (Degree) – Number of people the respondent knows within the target population.
- Respondent's Coupon Number – Coupon number of the respondent.
- Respondent's Recruiting Coupon Numbers – Coupon numbers respondent used to recruit others.

# Female Sex Workers – Vietnam, 2004

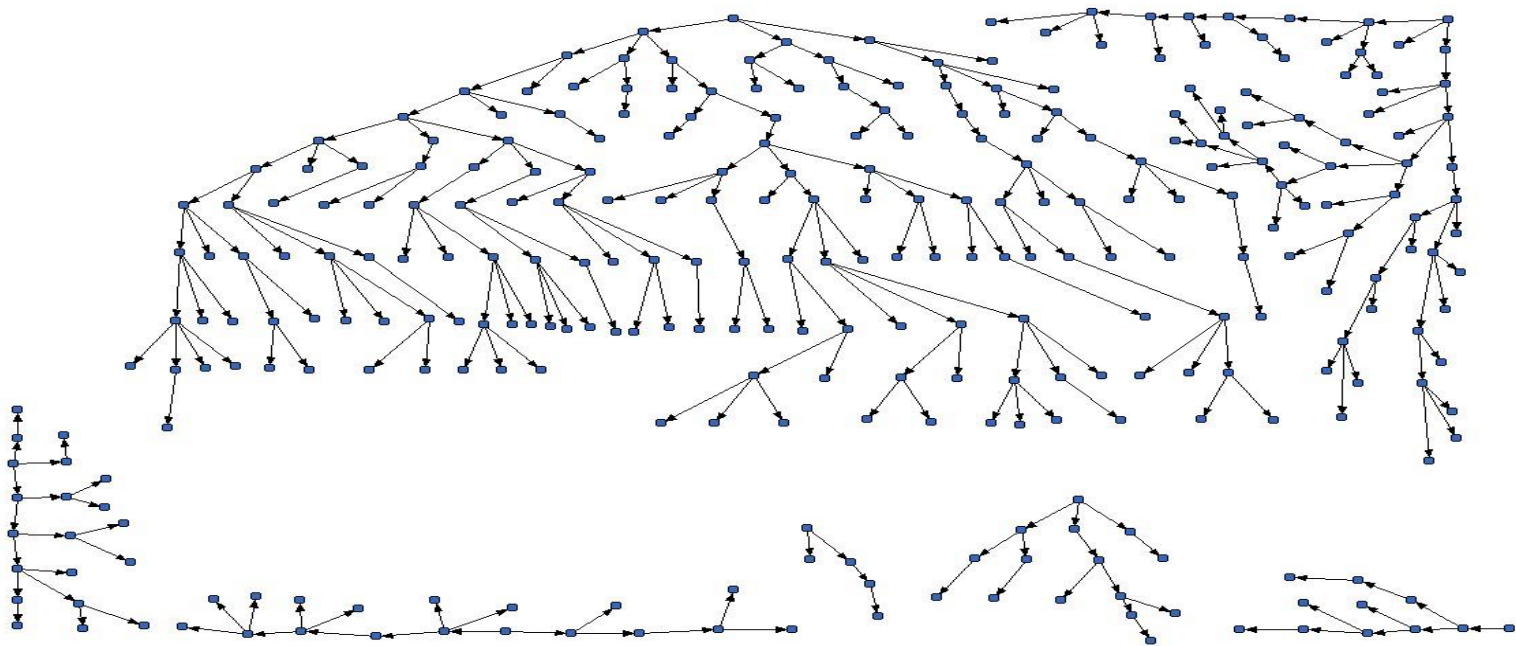


# Recruitment chain starting with a Gay seed (Seed #6, N=105), Dhaka, Bangladesh, 2006





# Respondent-driven sampling





## RDS Assumptions and Requirements

- Proportions will eventually reach equilibrium
- Connections are reciprocal
- Recruitment is occurring with same efficiency throughout the population
- The population from which a sample is gathered is infinitely large
- Participants' social network is sufficiently well connected

## RDS Assumptions and Requirements (cont.)

- Recruitment is non preferential
- Recruits are selected with probability proportional to their network size (recruiters with large network sizes are more likely to find someone to recruit)
- Recruits report their network size with accuracy sufficient

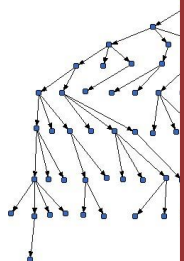
- ◉ <http://www.respondentdrivensampling.org/main.htm>
- ◉ RDSAT (statistical corrections)

● QUESTIONS?



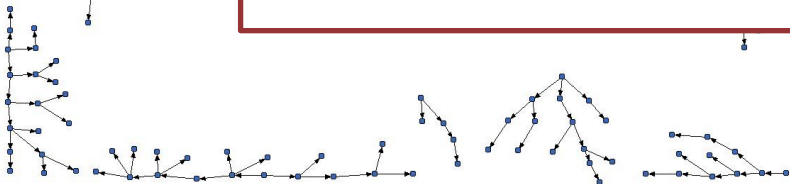
## Two approaches

### Chain-referral

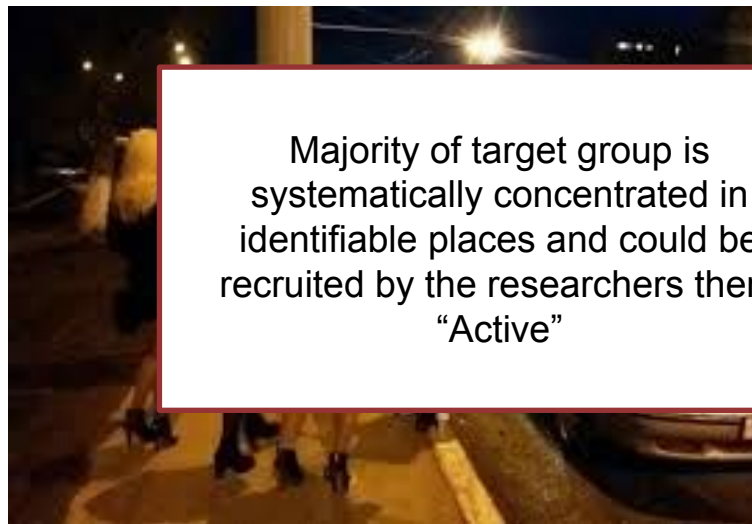


Members of the target group are well-connected and are willing to recruit each other

“Passive”



### Location



Majority of target group is systematically concentrated in identifiable places and could be recruited by the researchers there

“Active”



## Locations-based

---

- ⦿ Kind of cluster
- ⦿ Time–location Sampling (TLS)
- ⦿ Venue Day Time Sampling (VDT)
- ⦿ Temporal Spatial Sampling (TSS)
- ⦿ Time Venue Sampling (TVS)

# History

- Find and recruit from places  
(Watters & Biernaki (1989)), venue-based sampling  
□
- + estimation of N in each place (Carlson, Wang, Siegal, Falck, & Guo (1994) proportional quota sampling)  
□
- + randomization
- + adjusting for time

TLS (Lemp., et. al., 1994; MacKellar et al., 1996)

## **When to use**

- ⦿ Group is visible
- ⦿ Group is concentrated somewhere
- ⦿ We could get there
- ⦿ Absolute majority of this group use those places



- The **mapped “universe”** is inclusive of the diversity of the target population
- **Members of target population** have a chance of being sampled that is approximately known, equal, or can be adjusted for
- **Random selection** of venue, day, and time minimizes some of the biases of convenience sampling

- **Recruit eligible persons at VDT (variations):**
- –Consecutively
- –Systematically
- –Proportionately
- –Randomly

## **STEP1- Getting Started**

### **Understanding the Context**

- What is the **geographic area** of interest?
  - Is it the immediate city limits?
  - City and suburbs around it?
  - A larger jurisdiction?

## **STEP1- Getting Started Setting Goals and Objectives**

- Typical performance criteria to achieve a rigorous sample:
  - Data collection for no less than six months and no more than 12 months
  - Completing 14 sampling events per month
  - A minimum of 4 completed interviews per event
  - Completing 100% of sampling events
  - Complete  $\geq 90\%$  of the intercepts
  - Enroll  $\geq 75\%$  of the eligible men
  - Collect specimens with 80% of enrolled men
  - 500 subjects total (or calculated sample size)

## **STEP1- Getting Started**

### **Logistics and other considerations**

- **Logistics and other considerations**
  - *Biological testing*
  - *Survey instruments*
  - *Ethical considerations.*
  - *Institutional Review Board Approval*
  - *Reimbursement for time and effort / Incentives*
  - *Operations Manual*

## **STEP3-Formative Assessment / Community Buy In**

- Define the community of interest
- Ways of accessing the community
- The attributes of the community relevant to the specific public health issue
- Some of the tools of formative research
  - **Secondary Data Review**
  - **Focus Groups**

## **STEP4-Venue Univers Sampling Frame Construction**

- **Venue Identification (ID) Code**
  - Example : E = Social organizations , 1<sup>st</sup> -> ID CODE = E001
- **Venue Eligibility**
  - Any public or private locations attended by the priority population

### **Those excluded from monthly sampling frame :**

- Low levels of attendance of the priority population
- lack of safety
- disapproval by owners or managers

## **STEP4-Venue Univers**

### **Venue Identification**

- Venues and venue-day-time periods (VDTs)
- Elicitation of Socio-demographic Characteristics & Operational Barriers (Structural, Safety, Parking, Competing outreach activities)
- Collaboration with Venue Owners/Managers & Organizations
- Enumeration
  - Type I
  - Type II
  - Type III
- Attendance Levels



# Type I Enumeration

- Performed at **all** venues and is designed to capture :
  - that the venue is attended by the population
  - days and times of high attendance (VDTs)
  - estimates of how many people attend during these times
- Purpose : whether the venues gathered from formative research are actual venues that the MARP attends. ( **observation** )

# Type I Enumeration Form

## Behavioral Surveillance Type I Enumeration Form

Staff ID # \_\_\_\_\_ Date: \_\_\_\_\_

E Event #: \_\_\_\_\_ Venue Type ~~(1)~~\* \_\_\_\_\_ Clicker Count \_\_\_\_\_

Venue ID #: \_\_\_\_\_

Venue Name: \_\_\_\_\_

Begin-Time: \_\_\_\_:\_\_\_\_ a.m. p.m. End Time: \_\_\_\_:\_\_\_\_ a.m. p.m.

Day of the Week:      Sun   M   T   W   Th   F   Sat

\*Venue Type Codes: 1=Bar 2=Dance Club 3=Bus~~Estab.~~ 4=Health Club 5=Sex~~Estab.~~  
6=Social Org. 7=Street Location 8=Park 9=Other: \_\_\_\_\_

Comments (Weather, safety, etc.):

---

---

---

---

Draw area of intercept line/area below:

## Type II Enumeration

- Performed at **some** venues and is designed to capture :
  - Venue identifiers
  - enumeration counts
  - intercepts eliciting key information from patrons that establish membership in the priority MARP (e.g., gender, sexual behavior, IDU behavior)
  - whether intercepted persons are potentially eligible for the study
  - general sense of where and what kind of enumeration area is best for the venue
- **Purpose** : determine the number of eligible persons who attend a venue at a particular day and time period

# Type II Enumeration Form

**Behavioral Surveillance  
TYPE II ENUMERATION FORM**

Contact Person (Name, title, phone #)

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

Comments (Weather, safety, etc.):

Reviewed by:

Staff ID#: \_\_\_\_\_ Date: \_\_\_/\_\_\_/\_\_\_

Clicker #: \_\_\_\_\_ Date: \_\_\_/\_\_\_/\_\_\_

Supervisor #: \_\_\_\_\_ Date: \_\_\_/\_\_\_/\_\_\_

Year of Birth

Age if birthday is on or  
before today

Age if birthday is after  
today

Draw intercept line/area in box below.

CODES:					
<i>Ethnicity</i>					
<b>A = API</b>		<b>B = Black</b>	<b>H = Hispanic Latino</b>	<b>N = Native American</b>	
<b>W = White</b>			<b>U = Unknown</b>		
<i>County of Residence:</i>					
<b>SF = San Francisco</b>	<b>A = Alameda</b>	<b>CC = Contra Costa</b>	<b>M = Marin</b>	<b>SM = San Mateo</b>	<b>SC = Santa Clara</b>
<b>SCz = Santa Cruz</b>	<b>N = Napa</b>	<b>SA = Sonoma</b>	<b>SO = Solano</b>	<b>O = Other</b>	<b>U = Unknown</b>
<i>Language:</i>					
<b>E = English only</b>			<b>S = Spanish</b>		<b>B = Both equally</b>
<i>Sexual Behavior:</i>					
<b>M = Men Only</b>	<b>W = Women Only</b>	<b>B = Both Men &amp; Women</b>		<b>N = No Sex</b>	<b>U = Unknown</b>
<i>Willing to Participate? (Yes)</i>					

## **The criteria for including venues in the universe**

- The minimum effective yield is set at 8 individual during a four hour period
- Only VDTs that yield more than 75% of the target population

## Example of Venue Universe

Venue Info			VDT Information: January 2006							
ID	Type	Venue	M	T	W	R	F	Sa	Su	Current
P001	Park	Independence Park					2000-0000	2000-0000	1800-2200	Y
S001	Street	Wall and Gower					2000-0000	2000-0000	1800-2200	Y
B001	Bar	Echo Bar	1600-2000		2000-0000			2000-0000	1800-2200	Y
B002	Bar	Moe's	1700-2100	1700-2100	1700-2100	1700-2100	1700-2100	1700-2100	1700-2100	N
D001	Dance	T-Cozy					2000-0000	2000-0000	2000-0000	Y
P015	Park	Noe Valley Playground					2200-0200			Y

## STEP5-Random selection Sampling Calendar Creation

Venue Info			VDT Information: January 2006							
ID	Type	Venue	M	T	W	R	F	Sa	Su	Current
P001	Park	Independence Park					2000-0000	2000-0000	1800-2200	Y
S001	Street	Wall and Gower					2000-0000	2000-0000	1800-2200	Y
B001	Bar	Echo Bar	1600-2000		2000-0000			2000-0000	1800-2200	Y
B002	Bar	Moe's	1700-2100	1700-2100	1700-2100	1700-2100	1700-2100	1700-2100	1700-2100	N
D001	Dance	T-Cozy					2000-0000	2000-0000	2000-0000	Y
P015	Park	Noe Valley Playground					2200-0200			Y

# STEP5-Random selection

## Sampling Calendar Creation

- 1. Block out staff days off (e.g., holidays) \*
- 2. Schedule special events for the upcoming month (e.g., gay pride parade) or “oneoff” events

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27 Gay Pride 1000- 1400	28	29	30	31		



# STEP5-Random selection

Sampling

## 1. Independence Park

1. Friday 2000-0000
2. Saturday 2000-0000
3. Sunday 1800-2200

## 2. Echo Bar

1. Wed. 2000-0000
2. Saturday 2000-0000
3. Sunday 1800-2200
4. Monday 1600-2000

## 3. Wall and Gower

1. Friday 2000-0000
2. Saturday 2000-0000
3. Sunday 1800-2200
4. Wed 2000-0000

## 4. T Cozy Dance Club

1. Friday 2000-0000
2. Saturday 2000-0000
3. Sunday 2000-0000
4. Wed 2000-0000

## 5. Noe Valley Playground

1. Friday 2200-0200

# STEP5-Random selection

## Sampling Calendar Creation

### Primary sampling venues :

3. Randomly select, without replacement,  $n$  venues (typically 14-16) – determined when setting up performance criteria.
4. Arrange venues in order of least VDTs to most VDTs
5. Schedule least VDTs first moving through most VDTs
6. When venues has more than one VDT use dice (or other random selection method) to choose which VDT to schedule. Schedule the randomly chosen VDT on the first available day of the week.
7. Continue until all  $n$  (*e.g.*, 14-16) events are scheduled

### Alternative sampling venues :

8. For each event, group or list venues which have VDTs starting within the time period of the primary event
9. From this group choose, randomly without replacement, two alternate venues

# STEP5-Random selection

## Sampling Calendar Creation

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
		1	2 Echo Bar 2000-0000  A- Wall and Gower B- T-Cozy	3	4 Noe Valley Playground 2200-0200	5 T-Cozy 2000-0000  A- Wall and Gower B- Independence Park
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

## Goals

- TLS approximates **probability sampling method (Cluster Sampling)**.
  - Randomizing VDTs
  - Systematic sampling at the venue itself
  - The length of time spent in the field conducting sampling events

## **STEP5-Random selection**

### **Practical considerations for the sampling calendar**

- **Sampling Event Conflicts**
- **Canceling Events**
- **Alternates**
  - if there is a low traffic flow, staff must wait at least 30 minutes
- **Non-Random Events (Max 3 different Venues per month)**

### **Definitions of Traffic Flow ( in 15 min.)**

- Low flow = <20 clicked
- Medium flow = 21-50 clicked
- High flow = 50+ clicked

## **Step 6: Sampling Events / Recruitment**

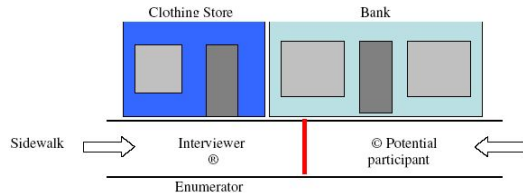
### **Key Activities during sampling events**

- **Enumeration** – count all persons who cross into a recruitment area
- **Intercepts** – approach and speak with designated persons
- **Eligibility** – ask person questions to determine whether they are able to participate
- **Enrollment** – encourage person to enter into the study
- **Complete survey** – take the participant through the entire survey
- **Counseling** – provide information about HIV/STDs and appropriate referrals
- **Specimen collection** – collect blood, oral or other fluid for HIV/STD tests

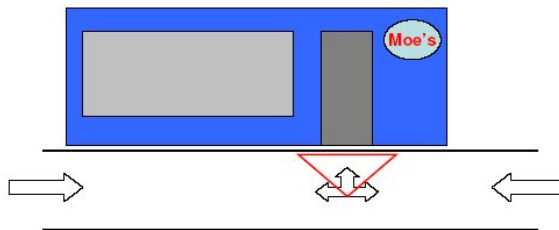
## Step 6: Sampling Events / Recruitment Systematic Sampling

- Enumerator counts every possible eligible person crossing intercept area
- Recruiters **systematically** approach enumerated persons
- Recruiter introduces study, assesses interest, determines eligibility, enrolls subject
- When all recruiters are occupied, enumerator continues to count
- When a recruiter is ready again, intercepts resume with the next person
- Enumerator can halt counting if problems arise
- Enumeration ends when the **four hour** time-period is complete.

## Step 6: Sampling Events / Recruitment Setting Up an Enumeration Area



© Christina / Keedra Luck, Flickr





## Step 6: Sampling Events / Recruitment

Strategies to successfully complete intercepts and enroll eligible subjects

### When not try to recruit

- Safety concerns
- People walking too fast (marathon walker)
- People on cellular/mobile phones, MP3 players
- Physical gestures / Body language
- Too high or drunk
- *Very firm refusal*
  - If possible, get their reason for refusal

## Step 6: Sampling Events / Recruitment Interview Options

<p><b>Gold Standard:</b> -Immediate interviews on site</p>	<ul style="list-style-type: none"> <li>• Best option</li> <li>• Interview follows recruitment</li> <li>• No loss to follow-up</li> <li>• Not always practical or desirable to respondents</li> </ul>
<p><b>Silver Standard:</b> —Same Day Appointments (SDA's)</p>	<p><b>Same Day Appointments</b></p> <ul style="list-style-type: none"> <li>• Used when interviewers are unavailable or respondent needs some time</li> <li>• Appointment is set for later in the field event</li> <li>• Enumerator tracks appointments to prevent conflicts</li> <li>• Potential loss to follow-up</li> </ul>
<p><b>Bronze Standard:</b> —Other Day Appointments (ODA's) —Last Ditch Efforts (LDE's)</p> <ul style="list-style-type: none"> <li>• All help minimize participation bias</li> <li>• By having some information on these potential subjects its possible to understand the direction of potential bias</li> </ul>	<p><b>Other Day Appointments</b></p> <ul style="list-style-type: none"> <li>• Used when time or conditions aren't conducive to immediate interview</li> <li>• Scheduled for the office or another location on another date</li> <li>• Contact information obtained, if accepted</li> <li>• Potential loss to follow-up, esp. without contact information</li> </ul>

## Step 12: Analysis

# Weighted Analysis

- TLS is held to approximate random sampling in that each venue/VDT has an equal chance of inclusion
- Enough venues/VDTs sampled
- Weighting has not often been used
  - venues have shown high heterogeneity of attendees
  - key outcomes were not found to be associated with venues
  - TLS usually produces many small clusters rather than a few large homogenous clusters □ minimize design effects

## Step 12: Analysis

# Weighted Analysis

$$\approx \frac{\text{Number of persons enrolled}}{\text{Number of eligible persons}}$$

- **Probability Weight** : Weighting can be achieved by using the enumeration count of each event as the basis for the weight.
- the ratio of the number of persons enrolled to the number of eligible persons at each recruitment event.

## Step 12: Analysis

# Weighted Analysis

1	2	3	4	5	6
Event	Enumeration	p of	Interview	p of	
ID	Count	enumeration	Count	interviews	P WEIGHT
1	16	0.007233273	1	0.0066667	1.08499096
2	107	0.048372514	19	0.1266667	0.38188826
3	67	0.030289331	11	0.0733333	0.41303633
4	60	0.027124774	10	0.0666667	0.40687161
5	913	0.412748644	20	0.1333333	3.09561483
6	353	0.159584087	15	0.1	1.59584087
7	102	0.046112116	22	0.1466667	0.31440079
8	65	0.029385172	10	0.0666667	0.44077758
9	132	0.059674503	23	0.1533333	0.38918154
10	397	0.179475588	19	0.1266667	1.41691253
<b>Totals</b>					
28	2212	1	150	1	9.5395153

## Step 12: Analysis

### Cluster Analysis / Stratified Analysis

- **Adjustment for Clustering** - Statistical software provides these adjustments by designating the venue as the group or cluster.
- a venue is differently attended by very different members of the MARP during specific day time periods could be counted as a **separate cluster**

- **Internal validity** strengthened by:
  - –High participation rate (performance goal: >75%)
  - –High eligibility assessment rate (>90%)
  - –High completion of VDT sampling events (>95%)
  - –No interviewer selection allowed
- **Statistical analysis adjusts** for venue attendance pattern (relative representation) and homogeneity within the venue (cluster)
- **External validity** depends on good formative research

- Typical performance criteria to achieve a rigorous sample:
- –Data collection for no less than six months and no more than 12 months
- –Completing 14 sampling events per month
- –A minimum of 4 completed interviews per event
- –Completing 100% of sampling events
- –Complete  $\geq 90\%$  of the intercepts
- –Enroll  $\geq 75\%$  of the eligible men
- –Collect specimens with 80% of enrolled men
- –500 subjects total (or calculated sample size)