

This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike License](https://creativecommons.org/licenses/by-nc-sa/4.0/). Your use of this material constitutes acceptance of that license and the conditions of use of materials on this site.



Copyright 2006, The Johns Hopkins University and Ron Gray. All rights reserved. Use of these materials permitted only in accordance with license rights granted. Materials provided "AS IS"; no representations or warranties provided. User assumes all responsibility for use, and all liability related thereto, and must independently review all materials for accuracy and efficacy. May contain materials owned by others. User is responsible for obtaining permissions for use from third parties as needed.

---

# Randomized Trials

---

Ronald Gray

---

# Topics

- What are randomized controlled trials (RCTs)?
  - Why do randomized trials?
  - Types of trials
    - Individual versus community RCTs
    - Phase 1-4 trials
  - Methodological issues
    - Design of trials (number and type of comparisons)
    - Sample size
    - Eligibility and enrollment
    - Consent
    - Randomization
    - Follow up
    - Endpoints
    - Analyses
    - Stopping rules
  - Ethical considerations
  - Safety monitoring
  - Case studies
-

---

# What are randomized controlled trials (RCTs)?

- A RCT is a planned experiment designed to assess the efficacy of an intervention in human beings by comparing the intervention to a control condition
  - The allocation to intervention or control is determined purely by chance (randomization)
  - RCTs are a subset of possible experimental designs
-

---

# Why randomize?

## ■ Randomization

- Avoids selection bias
  - Improves comparability of intervention and control arm populations (e.g., can match or block units of randomization)
  - Fulfills assumptions underlying tests for statistical inference
-

---

# What does “controlled” trials imply?

- “Controlled” implies predefined:
    - ❑ eligibility criteria
    - ❑ specified hypotheses
    - ❑ Primary and secondary endpoints (e.g., behavioral change, HIV incidence) to address hypotheses
    - ❑ Methods for enrollment and follow up
    - ❑ Rigorous monitoring
    - ❑ Analysis plans and stopping rules
-

---

# Why do RCTs?

- Observational and quasi-experimental designs are subject to potential bias and confounding due to:
    - Self selection (lack of comparability)
    - Observer bias
    - Secular trends (e.g., before and after study)
  - The RCT provides the “gold standard” for proof of concept
-

---

## Are results of RCTs always valid?

- RCTs can provide conflicting results
  - RCT design, execution or analyses can be flawed
  - Intervention vs control comparisons are internally valid, but restrictions on participant eligibility can reduce external validity (e.g., specific age or sex groups omitted)
-

---

# Types of Trials

## ■ Individually randomized trials

- Eligible individuals are randomized (conventional medical RCTs and many behavioral RCTs)
- Self-selection of persons volunteering for enrollment

## ■ Cluster randomized trials

- Clusters (e.g., communities, hospitals, or other aggregates of people (e.g., workplace, bars, brothels) are randomized, and all consenting persons enrolled
  - Less individual-level self-selection
-

---

# Why do Cluster-Randomized Trials?

- Nature of the intervention (e.g., mass media campaign, population-level interventions)
  - Acceptability and reduced stigma (everyone gets the same treatment within a cluster)
  - Can reduce participant self-selection → maximize generalizability
  - Can get data on many nested population subgroups
  - Allows assessment of population-level effects (Population Attributable Fraction)
-

---

# Limitations of Cluster-Randomized Trials?

- Cluster randomization more vulnerable to lack of comparability between study arms than individual randomization (fewer units of randomization, more correlated characteristics within members of clusters)
  - Cluster randomized trials increase sample size requirements and are less efficient than individual randomized trials due to intra-cluster correlation.
-

---

# FDA/WHO classification of individually randomized trials during course of drug or device testing:

- ❑ Phase 1, small sample size (<100), determine safety and preliminary evidence of effect
  - ❑ Phase 2, larger trial (100-200), assess safety, acceptability/tolerance, and probable effective dose
  - ❑ Phase 3 trials (500+), assess efficacy, acceptability, side effects and complications [Needed for FDA approval]
  - ❑ Phase 4: post-marketing trials in general population
-

---

# Trial designs

- Most trials have two arms (intervention vs control),
  - Multiple interventions can also be compared to a single control arm
  - Equivalency trials: head-to-head comparison of two or more treatments, without a control group (e.g., contraceptive trials)
  - Factorial designs e.g., intervention A, intervention B, intervention A+B vs control
-

---

# When should we do a trial?

- Trials must have:
    - a rationale based on prior observational data or biologic evidence
    - an explicit, testable and potentially falsifiable hypothesis
    - an uncertainty as to whether an intervention is efficacious (“ equipoise ”)
    - Reasonable expectation that benefits will exceed risk
    - An intervention that potentially can be randomized
-

---

# When is it unethical to randomize?

- Known effective treatment
    - Cannot use a placebo (e.g., trials to prevent mother-to-child HIV transmission). Need to provide standard of care
  - Personal choice
    - Cannot randomize very different interventions
    - For example, trials of different types of contraceptive (e.g., pill vs IUD), are ethically questionable because women have the right to select a method of their choice
    - (Can randomize within method type e.g., pill A vs pill B)
  - Risks of new treatment likely to exceed risks of existing treatment
-

---

# Sample size estimates for Individually randomized trials

- Most endpoints are measured as events in person time
    - Need to estimate person years (py) required to detect a specified rate in intervention vs control
    - Estimate sample size at enrollment and assumed loss to follow up to provide the person years needed
-

---

# Sample size estimation for individually randomized trials

- Specify type I and II error (e.g., power >80% to detect a difference significant at  $p < 0.05$ )
  - Specify an expected or meaningful difference in outcomes rates between intervention and control arms
  - Estimate losses to follow up which reduce person years
  - Estimate required sample size at enrollment using conventional formula (see last slides for methods)
-

---

# Sample Size in Cluster randomized trials

- Cluster randomized trials increase sample size requirements due to intra-cluster correlation.
  - Design effect (D) is the increase in sample size over individual randomization required to compensate for intra-cluster correlation (estimated by intra-cluster correlation coefficient or coefficient of variation).
  - To reduce sample size requirements try to select
    - More homogeneous clusters (less variability)
    - Larger cluster populations (lower coefficient of variation)
-

---

# Efficiency in cluster randomization

- Efficiency is increased with homogeneous population clusters
  - Efficiency increased by:
    - matching
    - stratified or blocked randomization
  - Larger population per cluster (reduces ICC or  $K$ )
  - There is a tradeoff between the number of clusters and the size of population per cluster
-

---

# Control groups

- Controls may receive no treatment (e.g., placebo) if there is no standard of care
  - If there is an established standard of care it would be unethical to withhold this from controls, so standard of care becomes the reference control
-

---

# Eligibility and Enrollment

- Eligibility is predefined to:
    - Ensure that participants meet the criteria for the intervention (e.g., have a specific disease for a therapeutic trial, are free of disease for a preventive trial etc.)
    - Usually eligibility is also defined by age, gender, race, state of health (absence of contraindications etc.)
    - The narrower the eligibility criteria, the less generalizable will be the results
    - Participants must consent to screening for eligibility
-

---

# Enrollment

- Enrollment occurs after:
    - Eligibility is established
    - Consent is provided after explanation of:
      - All study procedures
      - Risk and benefits
      - Measures to reduce risk
      - Voluntary participation (i.e., can refuse in part or in whole, or withdraw at any time)
      - Compensation for injury
      - Compensation for time and effort (e.g., money/gifts)
-

---

# Randomization

- Must be purely by chance
    - Random numbers
    - Random computerized allocation
    - Cannot randomize on any systematic characteristic (e.g., digits of SS#, attendance at a clinic etc.)
  
  - To maintain balance between arms one can
    - Individually match (e.g., by age, sex etc.)
    - Block randomization (e.g., randomize within blocks of 10-20)
-

---

# Follow up

- Follow up is conducted at predetermined intervals needed to detect the occurrence of trial endpoints
  - The frequency and duration of follow up will depend on:
    - Type of endpoint (e.g., response to treatment, development of new disease, progression of disease, behavioral change, sustainability of change)
    - The level of risk (higher the risk, more frequent the follow up)
-

---

# Loss to follow up

- Losses to follow up (LFU) must be minimized because:
    - Losses are often selective (e.g., high risk persons drop out of trials) and this introduces bias
    - Losses to follow up should be comparable in the intervention and control arms to avoid biased comparisons
    - Losses to follow up reduce study power by reducing the person-time of observation
-

---

# Blinding

- Blinding is done to minimize participant or observer bias
    - Double blinding (neither observer or participant know the arm of randomization)
    - Single blinding (observer but not participant knows the arm of randomization, e.g., cluster-level trials)
    - Unblinded (cannot conceal randomization, e.g., surgical interventions)
-

---

# Analyses

- Intent-to-treat

- Analyze all persons randomized, even if some do not receive the intervention or drop out before completion of treatment.
- Least biased and most conservative

- As treated (“per protocol”)

- Analyze only those who actually complete the treatment
  - Potentially biased by selection of the most compliant and often lowest risk population
-

---

# Statistical methods

- Outcomes at fixed points in time
    - Proportion with outcome at each follow up
    - Logistic or log binomial regression
      - Odds ratio (OR) or prevalence rate ratio (PRR) intervention/control
  - Events in person time
    - Rate of outcomes per 100 person years
      - Poisson regression, incidence rate ratio (IRR) intervention/control
  - Time-to-event
    - Time from enrollment until outcome
    - Cox proportional hazard regression, hazards ratio (HR)
    - Kaplan-Meier survival analyses, log rank test
-

---

# Stopping rules

- Trials should have predefined stopping rules to avoid:
    - Preventing undue risk to participants (e.g., treatment causes adverse effects)
    - Depriving the control group of an effective intervention
    - Continuing an ineffective intervention (“futility” or conditional power analysis)
-

---

# Trial Monitoring

- Trials must be approved by and monitored by Institutional Review Boards (IRBs) for ethics and safety
  - Trials should have an independent monitoring system to periodically review data and ensure participant safety
  - Data monitoring committee (DMC) or Data Safety and Monitoring Board (DSMB)
  - The DMC or DSMB should have the authority to terminate or change the trial procedures
  - Trials should report all adverse events, especially serious adverse events, and unexpected events
  - Complex regulations (Good Clinical Practice, GCP)
-

---

# Ethical principles (Belmont report)

- **Autonomy and respect for persons:**
    - Free and independent choice without coercion
    - Provision of informed consent
  - **Beneficence:**
    - Maximize benefit and minimize harm
  - **Justice:**
    - Equal opportunity to enjoy benefits
    - Provision of beneficial treatments to the population (social justice)
-

---

# Coercive inducement and full disclosure

- Participants should not be coerced by:
    - Denying treatment or benefits to persons who refuse (i.e., there should be some alternative treatment available)
    - By providing excessive compensation for participation (i.e., money or gifts)
    - By applying pressure to participate
  - There must be full disclosure of:
    - Reason for doing the trial, reason a person was selected for participation, who is funding the trial
    - Procedures entailed (eligibility, randomization, treatments)
    - Risks and benefits and measures taken to reduce risks
    - Confidentiality assurances
-

---

# Examples of RCTs

- Hormone replacement therapy (HRT)
  - STD control for HIV prevention
  - Behavioral interventions
  - Microbicides
-

---

# Postmenopausal Hormone Replacement Therapy (HRT) and Cardiovascular Disease (CVD)

---

Trials trump observational studies

# Observational studies of HRT and CVD

- Numerous case-control and cohort studies suggested that use of postmenopausal estrogens reduced the risk of cardiovascular disease (CVD) and of death from CVD
    - RR ~ 0.5.
    - Risks lower with higher dose estrogens.
  - Drug companies promoted HRT for “cardiac protection”
-

---

# Women's Health Initiative (WHIS) Study

(*JAMA* 2002;288:321)

- 16,608 healthy women 50-75
  - 
  - Randomized to conjugated estrogens + medroxyprogesterone acetate vs placebo
  - Follow up 5.2 years
  - Trial stopped due to increased risk and lack of net benefit
-

# WHIS Study Cardiovascular Disease Results

Outcome	IRR (95%CI)
Coronary heart disease (CHD)	1.29 (1.02-1.63)
Stroke	1.41 (1.07-1.85)
Pulmonary embolism	2.13 (1.39-3.25)
All Cardiovascular disease (CVD)	1.22 (1.09-1.36)

Other trials in women with pre-existing CVD showed similar increased risks

---

# Why did the randomized trial contradict the observational studies?

- **Self-selection** (i.e. healthier women, higher SES or higher educational status adopt supplements, as part of a “health conscious life style”).
  - **Physician prescribing habits** (avoid HRT in high risk patients).
  - **Duration** of observation and duration of estrogen use was often limited
  - **Age:** Women often relatively young (<60 years).
-

---

# Meta-analysis of observational studies of hormone replacement therapy and coronary heart disease

- No adjustment for socioeconomic status shows reduced risk
  - Adjustment for socioeconomic status shows no benefit
  - Source: Von Elm E, Eggers M. The scandal of poor epidemiological research. BMJ 2004;329:868-9.
-

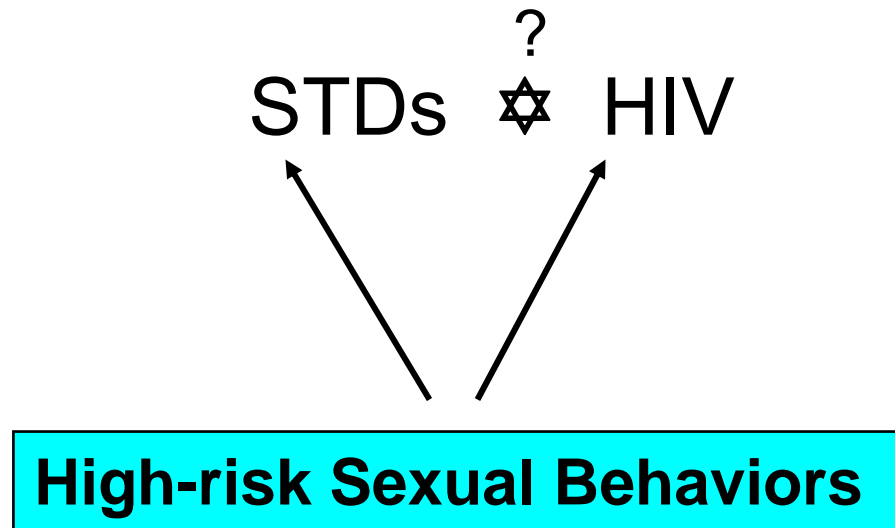
---

## Example of Community-randomized trials of STD Control for HIV Prevention

- Numerous observational studies suggest that treatable STDs are associated with HIV acquisition



# Problem of Confounding by Sexual Behavior



---

# Biological rationale for STDs and HIV Acquisition

- Genital ulcer disease (GUD) breaches mucosal barrier
  - Recruitment of HIV target cells and increased HIV receptor expression per target cell associated with inflammation
  - Elevated vaginal pH (eg, with BV or gonorrhoea) ☯  
HIV survival
-

---

# STDs Enhance HIV Transmission

- Increased HIV shedding from genital ulcer disease (GUD) or genital tract inflammation
- Disruption of mucosal barrier increases infectivity

Cohen. *Lancet*. 1998;351:5-7.

---

---

# HIV and STDs; public health policy

- Probable causal association between STDs and HIV acquisition/transmission at the **INDIVIDUAL LEVEL**
  - BUT
    - 1. Cannot resolve issue of behavioral confounding without a randomized trial.
    - 2. Public health question. Even if STDs increase individual risk, will STD control reduce HIV transmission/acquisition at the **POPULATION-LEVEL?**
    - Major policy question for HIV prevention in 1990s
-

---

# Community randomized Trials of STD Control for HIV Prevention

- Three trials:
    - Mwanza, Tanzania (Grosskurth et al Lancet 1995)
    - Rakai, Uganda (Wawer et al Lancet 1999)
    - Masaka, Uganda (Kamali et al Lancet 2003)
  - Tested the hypothesis that STD control can reduce HIV incidence
  - All 3 trials used community (cluster) randomization
-

---

# Why do Community-Randomized Trials of STD Control?

- **Nature of intervention:**

- STD control should ideally be community-wide to cover sexual networks, and maximize reduction of STD prevalence
  - Individual randomization would result in rapid STD reinfection since treated subjects would continue to have relations with untreated partners (**contamination**).
  - Providing STD treatment to both HIV+ and HIV- persons may reduce infectivity in HIV+ and lower HIV susceptibility in the HIV- (**Maximum bang for the buck.**)
-

---

# Strategies for STD Control and settings

- **Different strategies for STD control**
    - **Syndromic management** (Locations: Mwanza and Masaka)
      - Continuous access to services.
      - Only treat symptomatic STDs
    - **Mass presumptive treatment** (Location: Rakai).
      - Periodic treatment of all persons, every 10 months.
      - Treat asymptomatic +symptomatic STDs,
      - Maximize reduction in STD prevalence.
  - **Different HIV epidemic settings**
    - low grade epidemic Mwanza (HIV prevalence 4%)
    - mature generalized epidemics in Rakai (HIV 16%) and Masaka (HIV 12%)
-

# Community-based STD Control Trials

## Adjusted IRR (95% CI) HIV Incidence

	Rakai	Masaka A	Masaka B	Mwanza
HIV RR tmt/cont	0.97 (0.81–1.16)	0.94 (0.60-1.45)	1.00 (0.63-1.58)	0.62 (0.45-0.85)
	Mass Treatment	Syndromic management		

All trials showed reductions in treatable STDs  
Only one (Mwanza) showed an effect on HIV

# Kenyan trial of STD control for HIV prevention in commercial sex workers *Kaul et*

*al. JAMA 2004*

- Monthly presumptive (mass) STD treatment with azithromycin
- Reduced STDs
- No effect on HIV (IRR = 1.2, CI 0.6-2.5)

## Summary:

- Three out of four trials of STD control for HIV prevention show no effect on HIV incidence irrespective of the strategy for STD control
- Should STD control be promoted for HIV prevention?

---

# Behavioral intervention trials

- Observational data on behavioral interventions often problematic due to high degree of self-selection (e.g., persons accepting voluntary HIV counseling and testing (VCT), attending health education sessions etc.
  - Randomized trials of behavioral interventions are difficult because:
    - Hard to randomize (e.g., mass communications require cluster randomization)
    - Interventions often very intensive and demanding
    - **Response Bias:** Intervention may induce “desirable responses” (i.e., if participants educated to reduce risk behaviors, they may be less willing to admit such behaviors).
    - **Need biological end points** (e.g., STDs, HIV)
-

---

# Randomized trial of behavioral intervention in US minority women (Shain *NEJM* 1999)

- Mexican and African American women randomized to:
    - Intervention of 3 small-group sessions (3-4 hrs each) to recognize susceptibility, and acquire behavior change skills (n = 313),
    - Control received standard STD counseling (n = 304)
  - Followed up at 6 and 12 months to determine risk behaviors and STD infections (lab diagnosis)
-

# Results

	Intervention (%)	Control (%)
Follow up	84	80
STDs at 12 months	16.8**	26.9
2+ sex partners	32.5***	43.9
<5 unprotected sex acts	29.7***	20.2
5+ unprotected sex acts	70.3***	79.8

Intervention reduced reported risk behaviors and STDs

---

# Efficacy of Voluntary Counseling and testing

(VCT) Voluntary Counseling and Testing Efficacy Study Group,  
*Lancet* 2000)

- Randomized individuals and couples to:
    - VCT
    - Controls received basic health information
    - Followed up at ~6 and 12 months
    - Assess risk behaviors and STD infections
-

# Main results presented by authors

Decline in unprotected sex with non-primary partners, baseline to 12 months	VCT	No VCT
Males	↓ 35%**	↓ 13%
Females	↓ 39%**	↓ 17%

- Authors emphasized changes from baseline, rather than differentials between arms during follow up
- No data provided on STDs or HIV incidence (no effect)
- Can self-reported behaviors be used to assess efficacy of VCT? (response bias??)

# My interpretation of results

Unprotected sex with non-primary partners at 12 months	VCT	No VCT
Males	38.4% <sup>ns</sup>	38.4%
Females	22.3% <sup>ns</sup>	25.2%

- The prevalence of high risk behaviors at 12 months was similar in both arms! No effect
- This is the public health question of relevance
- Authors committed to VCT, so “salvaged” findings

---

# Meta-analyses of RCTs

- Many trials are conducted on varying populations with similar interventions
  - Need to combine data across trials using met-analysis
  - **CONSORT** agreement with all major medical journals to present results in a comparable format and allow pooling of data across trials
-

---

# Meta-analysis of trials of microbicides (Nonoxynol-9) for HIV and STD prevention (Wilkinson Lancet Infect Dis 2002)

- Meta-analysis of nine RCTs of N-9, all conducted in STD clinics or commercial sex workers
  - HIV incidence IRR = 1.12 (CI 0.88-1.42) [one trial found significantly increased HIV with N-9]
  - Gonorrhoea IRR = 0.91 (CI 0.7-1.4)
  - Genital ulcers (GUD) IRR = 1.18 (1.02-1.36)
  - Conclusion: N-9 increases genital ulceration and may increase HIV risks and does not reduce STDs
-

---

# Limitations to N-9 trial generalizability

- N-9 must be used with every sex act, all trial populations only included women who have frequent intercourse (e.g., STD clinics, CSWs)
  - Frequent use of N-9 causes irritation and micro-ulceration
  - Cannot determine whether women who have less frequent intercourse (e.g., general population) might benefit from N-9
  - Cannot ethically do a RCT in a general population because N-9 increased GUD, and might increase HIV in high risk populations
  - Poor choice of study populations for RCTs may have eliminated a potentially useful microbicide (i.e. eligibility criteria were inappropriate)
-

---

# Operations research can be a randomized trial

- Operations research of family planning outreach versus standard services, Rakai
  - Community randomized trial
  - Modern method use at follow up
    - Intervention 21.3%
    - Control 16.4% ( $p = 0.001$ )
-

---

# Is there a publication bias?

- Many trials sponsored by pharmaceutical industry
    - Biased reporting of adverse events (e.g., Vioxx)
  - Drug companies tend to publish positive trials, but often do not publish negative or equivocal trials
  - New trial registry:
    - All trials must be registered at initiation, and no unregistered trials will be published
    - Register of trials provides an open source data base
-

---

# Do trials affect policy?

- Trials often have a major impact on policy:
    - RCT of IUDs: Lippes loop vs Copper T, showed lower pregnancy and complication rates with copper devices, and change policy throughout the world
    - WHIS study showed HRT risks > benefits, reduced prescribing
    - Cox-2 inhibitors (Vioxx, celebrex etc.) increased CVD, reduced prescription
    - STD control for HIV prevention (3/4 trials showed no effect), but policy on STD control is unchanged [old habits die hard]
-

---

# Conclusions

- RCTs are the “gold standard” for proof of efficacy
  - Trials trump all other forms of evaluation but:
    - RCTs are imperfect, and trials may contradict one another
    - RCTs may not be generalizable to a broader population (e.g., N-9)
    - Some questions are not amenable to RCTs.
-