An Implementation Science Evaluation Agenda

Bridging the gap between evaluation and implementation
Outline

• Overview of IS
• Identify the implementation cascade
• Methodological components
• Impact Evaluation
• IE study designs
What is IS?

• Methods to improve the uptake, implementation, and translation of research findings into routine and common practices.

• Maintains rigor of traditional biomedical research with better external validity
  – Proof of concept ➔ Effectiveness
  – Mid-course corrections permitted
Overall Goals of an Implementation Science Agenda:

- Choosing interventions strategically
- Focusing them where they will have maximum benefit
- Improve implementation efficiency
  - Better management
  - Strategic integration with other services
- Maximize long-term benefit, not results for the annual report
I S Approach focuses on how we:

• Transfer, adapt and scale up interventions from one setting or population to another
• Assess whether interventions are efficient and effective by conduct of robust evaluations at scale
• Make informed, evidence-based choices between competing:
  – components within a combination strategy
  – strategies for delivery
Outcomes of IS

- Effectiveness
- Optimal delivery of services (Value for the money; efficiency)
  - Cost effectiveness
  - Most efficient strategies for implementation
IS research begins with identifying the Implementation Cascade

An intervention can be developed to plug up any point of leakage
PMTCT Cascade: Most Critical Thing for PMTCT is Number of Women Completing Cascade

P. Barker, WHO Mtg Nov 2008

100 HIV+ mothers

Attend ANC clinic 92%

Counseled and tested for HIV, CD4 75%

Get ARVs (pre-and perinatal) 50%

CD4 >200 / CD4 <200

sdNVP / HAART (8% MTCT): 3 infected
AZT/sdNVP/ HAART (3% MTCT): 1 infected
HAART all (2% MTCT): 0.6 infected

Missed - no PMTCT

Overall Program Effectiveness (early MTCT)

sdNVP alone: 22.5% tx
sdNVP +ART: 19.5% tx
AZT/sdNVP: 17.5% tx
HAART: 17.1% tx

No ARV (25% MTCT): 16.5 infected
Challenges in Achieving Promise of ART for Prevention: Individual factors

- **Unaware of HIV Status**
  - Unaware of individual risk for HIV
  - Vulnerable and disenfranchised populations at higher risk
  - Limited access to HIV testing
  - Refusal of testing due to denial or stigma
  - Provider practice of risk-based versus routine screening

- **Late Diagnosis of HIV Disease**
  - HIV testing not accessible or underutilized
  - Providers and individuals unaware of HIV symptoms or risk

- **Failure of HIV Positives to Link to Care**
  - Limited access to care services
  - Barriers (e.g., substance use, homelessness, mental illness)
  - Stigma (e.g., racial/ethnic, gender)
  - Mistrust of the healthcare system

- **Late Initiation of ART**
  - Limited access to treatment/waiting lists for ADAP
  - Limited treatment literacy and fear of side effects
  - Provider attitudes regarding ART initiation for some patients

- **Inability to Maintain Viral Suppression**
  - Limited access to effective antiretroviral treatment
  - Barriers to adherence (e.g., substance use, homelessness, mental illness)
  - Limited availability of adherence support
Challenges in Achieving Promise of ART for Prevention: Health system facilities

- **Bricks, mortar and commodities**
  - Central and decentralized facilities
  - Test kits, drugs

- **Human resources**
  - Task shifting
  - Training

- **Supply side incentives**
  - P4P
  - COD
  - CCT

- **Maintain existing services**
  - Build on existing platforms while not sacrificing quality
Specific IS example questions

• What are the best ways to optimize service delivery:
  – Balance of fixed and mobile clinics
  – Most efficient and effective methods to accelerative task shifting
  – Effectiveness and efficiency of vertical versus integrated services

• How can we improve access to programs?
  – Whom to target (optimal time) to begin treatment for HIV and TB to maximize clinical and public health benefits
  – Best methods to
    • Identify those who are eligible (on-going screening)
    • Decentralize quality care
    • Increase adherence and retention
Specific IS example questions cont

• What are the most robust methods to optimize or amplify the impact of prevention?
  – incentives or economic opportunities
  – Optimal combination of strategies to enhance ARV-based prevention

• How can we adapt the health platform PEPFAR helped develop to strengthen delivery system for other health outcomes
  – TB
  – MCH, Reproductive health
Shifting Evaluation Paradigm

• From retrospective, external, divorced evaluation
  – Top down
  – Approximate whether program worked or not

• To prospective, internal, and operationally driven impact evaluation /externally validated
  – Set program learning agenda bottom up
  – Consider plausible implementation alternatives
  – Test scientifically and adopt best; what works
  – Just-in-time advice to improve effectiveness of program over time
Fundamental Methodological Components of Implementation Science

- Monitoring and Evaluation
- Operations Research
- Modeling
- Impact Evaluation
Monitoring and evaluation

- Fidelity: whether the program being implemented as designed and planned
- Program components: whether inputs are sufficient to achieve the desired outcomes
- Reach and targeting: whether the program benefits getting to the intended recipients
- Process indicators and intermediate effects: whether expected program outputs and outcomes moving in the right direction?
Traditional M & E

- Is the program being implemented as designed?
  - Could the operations be more efficient?
  - Are the benefits getting to those intended?
  - Costing and accounting
- Monitoring trends
  - Are indicators moving in the right direction?
- No ability for intervention attribution
Operations Research

• Identify the implementation/delivery challenges and bottlenecks of a particular program:
  – Supply chain management
  – Health systems, clinical, and laboratory infrastructure
  – Efficiency and effectiveness of outreach and retention
  – Information systems

• Assess the optimal allocation of resources for the program
Impact evaluation = Smart Implementation

Taking advantage of rolling out programs at scale; learning by doing: integrate evaluation with implementation
What does impact mean?

• The word impact is often misused as a synonym for higher-level outcome

• Most simply, impact means “the effect or influence of something on something else”

• In IE, impact is the portion of an observed change in an outcome caused by the intervention of interest
What is Impact Evaluation?

Counterfactual analysis to single out the causal effect of an intervention on an outcome

- Compare same individual with & without “something” at the same point in time
- Compare the “factual”—the observed outcome among those who are “treated” with the counterfactual among a control or comparison group

Counterfactual Criteria

- Treated & counterfactual groups have identical initial average characteristics
- Only reason for the difference in outcomes is the intervention (treatment)
Counterfactuals control for extraneous factors that cause outcomes to change anyway

• Selection Bias
  • Participants who voluntarily participate may have different risk profiles than those who don’t

• Endogenous Change
  – Secular changes or drift (long term trends in community, region or country)
  – Maturational Trends (Individual change)
  – Interfering Events

• Known and unknown confounders
• Hawthorne/cohort Effect
Monitoring vs Impact evaluation

- Change over time
- Compare results before and after on the impact of treatment
- Monitoring measures change; IE measures impact
Problem with poor comparisons

• Before and After
  – Does not control for endogenous factors
  – Cannot control for Hawthorne effect

• Enrolled versus not enrolled
  – Participants who voluntarily participate may have different risk profiles than those who don’t

• Time trends with no counterfactual
  – Spurious associations
"I'm beginning to think it's salsa that's causing your mood swings."
Results Chain

INPUTS
- Financial, human, and other resources mobilized to support activities.

ACTIVITIES
- Actions taken or work performed to convert inputs into specific outputs.

OUTPUTS
- Products resulting from converting inputs into tangible outputs.

OUTCOMES
- Use of outputs by targeted population.
- The final objective of the program.
- Long-term goals.

FINAL OUTCOMES
- Not fully under the control of implementing agency.
- Changes in outcomes with multiple drivers.

Implementation (SUPPLY SIDE)
- Budgets, staffing, other available resources.
- Series of activities undertaken to produce goods and services.

Results (DEMAND+SUPPLY)
- Goods and services produced and delivered, under the control of the implementing agency.
Monitoring and Impact Evaluation

- monitoring to track implementation efficiency and fidelity (input-output)
- impact evaluation to measure effectiveness (output-outcome)

INPUTS

Fidelity

OUTPUTS

EVALUATE EFFECTIVENESS

$\text{BEHAVIOR}$

$\text{MONITOR EFFICIENCY}$

OUTCOMES

$\text{EFFECTIVENESS}$

$\text{MONITOR}$
Impact Evaluation Answers...

• What was the effect of the program on outcomes? (*causal attribution*)

• How much better off are the beneficiaries as a result of the program? (impact/strategies for implementation)

• What happened compared to what would have happened without the program

• Is the program cost-effective?
  
  – Depends on costing, plus robust measures of effectiveness as determined from IE
IE methods of evaluation
Study Designs/Analysis for IE
Design and analysis techniques overlap

Epidemiology

1. RCTs
   (indiv. or cluster)

2. Cohorts

3. Case-control
Study Designs/Analysis for IE

Design and analysis techniques overlap

Epidemiology

1. RCTs (indiv. or cluster)
2. Cohorts
3. Case-control

Impact Evaluation

1. Experimental designs (LST, cluster and other randomized approaches)
2. Quasi-experimental:
   • Matching
   • Difference-in-difference
   • Discontinuity designs
Randomized Approaches

• The “gold standard” in impact evaluation

• Gives each eligible unit the same chance of receiving treatment
  – Lottery for who receives benefit
  – Lottery for who receives benefit first

• Ensures comparability between those who did and did not receive the intervention

• Ensures that impact measurements are not confounded by extraneous confounding factors
  • Permits demonstration of mitigation of a negative trend
Other rationales for randomization

• Universe of eligible individuals is typically larger than available resources at a single point in time

• Fairest and most transparent way to assign benefit may be to give all an equal chance of participating → randomization

• Permits unequivocal, simple answers that inform policy and stand up to political and scientific queries
As opposed to traditional randomized “efficacy” controlled trials (RCTs)

• Targets individuals or communities
• Evaluator directly controls implementation
• Intensive in terms of cost, training
• Implementation is “perfect”
• Long duration
• Does not consider adherence, migration or other factors that might have diluted the effect
• Highest internal validity (not real world)
Efficacy vs. Effectiveness

Efficacy – the effect under ideal circumstances (i.e. laboratory, clinical trials; proof of concept): **Narrow and Deep**

Effectiveness – the population effect under real world circumstances / the effect at scale; aka **Impact Evaluation: Broad and less Deep, but still rigorous**
Large Simple Trials or Interventions under "real world" conditions of use

- typically include a large number (several thousand) people and
- broad eligibility criteria,
- simple enrollment procedures
- LSTs collect minimal data, and use clearly defined, easy to assess outcomes as endpoints
- Less frequent follow up.

Peto, R; Collins, R. Gray, R. J Clinical Epi, 1995
When randomization is not appropriate or possible

- Ethical concerns
- Interventions of known effectiveness
- Full coverage programs
- Political will
- Deciding who is covered can be politically and ethically tricky
Quasi-Experimental Designs and Analytic techniques (1)

- Generates a valid counterfactual without randomly allocating the program or policy
- Reduce variance and restricts comparisons to comparable subsets
- Use existing data when possible and are often quicker and cheaper than experimental designs
- Can be more analytically intensive, may suffer from selection bias, and violate many statistical assumptions

Quasi-Experimental Designs and Analytic techniques (2)

• Whether the intervention and comparison group are comparable will *always* be arguable

• The comparability between those who do and do not receive the intervention achieved through randomization is approximated
  • Need to understand who was exposed and why

• Various types of comparison studies
  – Cohort (prospective): Valid measure of incidence
  – Cross-sectional: Hawthorne effect less likely
Matching (at design or analyses)

• Matches the treatment group from a larger survey or cohort based on exposure.
• Matches are selected on the basis of similarities in observed characteristics.
• Assumes no selection bias based on unobservable characteristics.
• Examples include: *Propensity score matching* (2 groups) *Marginal Structural models* (*MSMs*) within one observational cohort; *Matched community designs*

Source: Martin Ravallion
Adaptive Designs (common sense RCTs)

• Flexible designs that permit mid-trial modifications without compromising the ultimate statistical assessment of results.

  – Traditionally focused on early termination rules
  – Subsequent focus on sample size modifications, dose finding and modifying the intervention or its intensity

Example: Drop-the-Loser

Interim results suggest that some arms are inferior and are dropped from the study.

Hybrid, nested designs and analyses are possible!

- Smaller, representative cohort embedded in a larger repeat cross-sectional evaluation;
  - Nationally representative cohort
  - Cohort in certain regions
- Concurrent or additional experiments within quasi-experiments
- Qualitative, in-depth studies
Modeled counterfactuals

• Analytic methods that infer causality from observational designs
• Epidemiological Modeling
Choosing Your Methods

Identify the “best” possible design given the context

• Best design = fewest risks for error

• Is the result valid for “everyone”?  
  – External validity  
  – Proof of concept versus global treatment effect
Consider Randomization First

• Minimizes selection bias
• Balances known and unknown confounders
• Most efficient (smaller Ns)
• Simpler analyses
• Transparency
• Decision makers understand (and believe) the results
3ie

- International initiative for impact evaluation
- Conferences
- Workshops
- Systematic Reviews
- TA
Challenges unique to prevention

• No surrogate outcome
• Long time horizon
• Large sample size
• Control group vs standard of care
  – Short and long-term outcomes and ethics
Problems with High-Risk Behavior as a Surrogate Outcome for HIV

• Self-report is unreliable and inconsistently associated with HIV infection or other biological markers of unprotected sex
  – Recall bias
  – Social desirability bias
  – Unknown partner risk (↓ risk behavior w/ ↑ risk partners)
    • Problem of risky sex with low risk partners, low risk sex (protected) with risky partners

• No formula to translate high-risk behavior into infections averted & the effect is variable

• Necessary but not sufficient for transmission
ACASI vs. Face-to-Face Interview

910 women enrolled and randomized

PSA testing +
FTFI
(N = 460)

PSA testing +
ACASI
(N = 450)

• 196 participants (21.5%) had biological evidence of recent semen exposure (i.e. tested positive for PSA)
  - 104 participants in ACASI arm
  - 92 participants in FTFI arm
## Discordant PSA Results and Self-Reported Behaviors

<table>
<thead>
<tr>
<th>Reported activity past 2 days</th>
<th>Total PSA + (N=196)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
</tr>
<tr>
<td>No sex</td>
<td>23</td>
</tr>
<tr>
<td>Only condom-protected sex</td>
<td>71</td>
</tr>
<tr>
<td>Total</td>
<td>94</td>
</tr>
</tbody>
</table>
Problems with Biological Outcomes as Interchangeable Surrogates

Causal chains: “exposure” and “susceptibility” vary for each outcome