Evaluating Complex HIV Programs: A tale of two studies

World Bank June 2011

Audrey Pettifor, PhD MPH
Department of Epidemiology, University of North Carolina
Overview

- How to evaluate complex HIV prevention programs
- Two examples:
  - loveLife
    - Designs
    - Choices made
    - Attribution/Causality
    - Lessons Learned
  - HPTN 068
    - Design
    - Contamination/Spill-over
HIV Prevention

- Globally more than 34 million people living with HIV/AIDS at the end of 2010
- The causes of HIV are multifaceted -- this requires Combination Prevention
- Interventions that must be taken to scale
HIV Prevalence by Age and Gender among South African youth age 15-24 years

How to evaluate complex Interventions?

- **RCT— gold standard**
  - Cluster randomized
    - Step Wedge
    - Parallel
  - Individual randomization

- **Quasi-Experimental**
  - Control Group? Randomization?

- **Observational**
  - Pre-Post
  - Cohort
  - Population surveillance

- **Mathematical Modeling**
  - Extrapolate from existing data
What is loveLife?

A national HIV prevention initiative for young people combining high-powered multi-media with comprehensive youth-friendly sexual health services in public clinics, and countrywide community outreach and support programs.
To halve the rate of new HIV infections, and significantly reduce sexually transmitted disease and pregnancy among 15-20 year olds.
**OUTCOMES**

- To provide sexual health education and care for adolescents in a non-clinical environment.
- To build the capacity of local NGOs/CBOs to address the needs of young people in communities with regards to HIV prevention.
- To improve the quality and accessibility of reproductive health services delivered to adolescents at local clinics.
- To provide personalized reproductive health information as well as counselling services.

**FACILITIES**

- **Y-Centres:**
  Multipurpose recreational venues for young people.
- **Franchises:**
  Community NGOs/CBOs endorsing the loveLife approach and implementing loveLife programmes.
- **NAFCI Clinics:**
  Clinics that adopted the National Adolescent Friendly Clinic Initiative.
- **loveLife Call Centre**
  Helplines for young people (theJunction) and their parents (Parentline).

**Positive Lifestyle EXPERIENCE**

Increasing the number of young people reached by loveLife’s programmes and activities by taking the loveLife message/programmes outside facilities to young people.

Regional bases for outreach activities by groundBREAKERS and mpointshis.

- **loveLife Games**
- **loveTrain**
- **loveTours**

*Although not a strictly a ‘facility’ the loveLife Call Centre provides a valuable service to young people and their parents.*
Challenges in evaluating loveLife

- Cannot randomize
  - Self-selection bias
- National program - in theory, everyone is exposed. Who is the control?
- No baseline
- Measuring exposure is complex
- Self-reported data
- Attribution? Many other HIV prevention programs in South Africa
How to evaluate loveLife

- Evaluation team had different opinions on how best to evaluate a national program
- This resulted in 2 approaches:
  - Nationally representative household survey, powered to detect a change in HIV over time
  - Quasi-experimental study, facility based—powered for change in HIV and STDs over time
- Lots of monitoring data
National survey: Methodology

- Nationally Representative Household survey
- Repeated, cross-sectional surveys
- Males and Females age 15-24 years
- Three stage disproportionate, stratified sample (EA, segment, youth in household)
- Interviewer administered questionnaire
- Oral fluid specimen collected using OraSure and tested for HIV-1/2 using Vironostika ELISA
- Conducted March-August 2003
- Final sample consisted of 11,904 young people

Quasi-Experimental Design

- Quasi-experimental community based, repeated, cross-sectional surveys
- 3 “arms”: Y-Centre (n=11 communities), NAFCI (n=11 communities), Comparison (n=11 communities)
- Households in EA enumerated, EA segmented and one segment randomly selected, one young person per household randomly selected
- Interviewer administered questionnaire
- Anonymous testing for HIV, NG, CT
- Baseline August 2002- Feb 2003
- No follow up ever funded

Framework for determining program impact

  - Adequacy - Did the expected change occur?
  - Plausibility – Did the program seem to have an effect above and beyond other external influences?
  - Probability – Did the program have an effect?
## Determining Plausibility of an effect (1)

<table>
<thead>
<tr>
<th></th>
<th>Tools to measure</th>
<th>Potential Indicators</th>
<th>Possible Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Provision</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are the services</td>
<td>Monitoring</td>
<td>*Number of different LL programs</td>
<td>*235 NAFCI clinics *532 Schools participating</td>
</tr>
<tr>
<td>available?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are they accessible?</td>
<td>Monitoring</td>
<td>*Hours of operation *Adolescent Friendly?</td>
<td>*Open afternoons / weekends, provided in school and CBOs *All programs adolescent friendly</td>
</tr>
<tr>
<td>Is the quality</td>
<td>NAFCI- Quality</td>
<td>*Quality Score</td>
<td>*35% of NAFCI clinics attained 90% of standards for adolescent quality care</td>
</tr>
<tr>
<td>satisfactory?</td>
<td>Assurance Assessments</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Utilization</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are the services</td>
<td>Monitoring</td>
<td>*Number of youth services provided</td>
<td>*153,543 youth participated in LL programs in Q4 2004 *108,570 clinical services provided to 10-19 yr olds in Q4 2004</td>
</tr>
<tr>
<td>being used?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Determining Plausibility of an effect (2)

<table>
<thead>
<tr>
<th>Coverage</th>
<th>Tools to measure</th>
<th>Potential Indicators</th>
<th>Possible Outcomes</th>
</tr>
</thead>
</table>
| Is the target population being reached? | Monitoring Evaluation | *Number in target population accessing services  
*Number of youth in HH survey reporting awareness of and exposure to loveLife programs  
*Ecological measures of program in area | *85% of youth reported being aware of the loveLife program  
*34% report participation |

<table>
<thead>
<tr>
<th>Impact</th>
<th>Tools to measure</th>
<th>Potential Indicators</th>
<th>Possible Outcomes</th>
</tr>
</thead>
</table>
| Improvements in disease patterns or health related behaviors? | *National HH survey of HIV and sexual behavior  
*Community Based survey | *HIV prevalence  
*Sexual Behaviors  
*Norms and attitudes with regard to HIV prevention | *Observed changes in HIV and behavior based on follow-up surveys |
Factors to strengthen plausibility of loveLife’s effect (1)

- Measure confounders - adjust in multivariable analyses*:
  - Males HIV infection OR 0.60 (95% CI 0.40-0.89)
  - Females HIV infection OR 0.61 (95% CI 0.43-0.85)

- Consistency of results across outcomes**?
  - Talk to parents about HIV
    - OR 1.59 (1.21-2.08)
  - Changed behavior due to HIV
    - OR 1.57 (1.16-2.12)
  - High sense of future optimism
    - OR 1.39 (1.16-1.66)

- Dose response associations*
  - HIV infection
    - One program (vs. 0) OR 0.68 (0.54-0.85)
    - Two programs (vs. 0) OR 0.44 (0.33-0.58)
  - Always used a condom w/last partner
    - One program (vs. 0) OR 1.47 (1.18-1.83)
    - Two programs (vs. 0) OR 2.28 (1.71-3.05)

*adjusted for completed high school or not, race, age, urban/rural residence, marriage, electricity in the home, participation in youth group, tested for HIV, knowing someone who died of AIDS, and awareness of two different national HIV prevention campaigns
Other methods to strengthen plausibility (2)

- Mathematical Modeling
- Cost effectiveness
- Spatial mapping
- Detailed program monitoring & process evaluation data
Lessons Learned

- Pre-intervention data was critical
  - Design evaluation BEFORE you implement the program—rigorous evaluation only designed and implemented in year 3 of the program—data only available in year 5.

- Good evaluations cost $$$
  - 10% of total program budget—minimum
  - Lack of “gold standard” evaluation data until 5 years into the program was problematic for continued program funding

- Measuring breadth and depth
  - Program ‘awareness’ vs ‘participation’

- If program implementation and evaluation had been designed in parallel—Step-Wedge type design could have perhaps been implemented—Can’t always go everywhere at the same time.
HPTN 068: SWA KOTEKA
(It Is Possible!)

Effects of Cash Transfers and Community Mobilization for the prevention of HIV in young South Africa women
Study Design

- RCT
- Factorial, Multi-level
  - Conditional Cash Transfer to poor young women and their families- randomize households (ages 13-20, grades 8-11)
  - Community mobilization- randomize villages in which young women live (n=22) (target men 18-35 yrs)
<table>
<thead>
<tr>
<th></th>
<th>CM Yes</th>
<th>CM No</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CCT Yes</strong></td>
<td>CCT+ CM</td>
<td>CCT Only</td>
</tr>
<tr>
<td><strong>CCT No</strong></td>
<td>CM only</td>
<td>No intervention</td>
</tr>
</tbody>
</table>
Primary Objectives

- **CCT**: To determine whether young women who are randomized to receive cash transfers conditional on school attendance have a lower incidence of HIV infection over time compared to young women who are not randomized to receive cash transfers.

- **CM**: To determine whether young people ages 18-35 living in villages that are randomized to a mobilization intervention focused on young men demonstrate positive changes in gender norms compared to young people living in villages that are not randomized to mobilization.

- **Interaction**: To determine whether young women who receive the CCT and live in communities that are mobilized have a lower incidence of HIV compared to young women who receive the CCT alone.
Study Site: Agincourt, South Africa
A view of the Agincourt HDSS Fieldsite Area. (K.Schatz 2003)

Copyright © 2000-2009 University of the Witwatersrand
Cash Transfer

- Randomize 2660 HIV negative girls (living in the 25 study villages) to get monthly cash transfer payment conditional on monthly school attendance

- Girls in grades 8, 9, 10 and 11 in Jan 2011, ages 13-20 yrs

- Transfer monthly, to female HH and girl
  - R300 per month based on 80% attendance at school
  - R200 to female HH and R100 to girl

- Total intervention time 3 years

- Assessments at baseline, 12, 24 and 36 months
Mobilization

- Target men 18-25
- Randomize 24 villages- half get community mobilization and half do not
- Working with Sonke Gender Justice to conduct outreach activities in the community that aim to mobilize the intervention communities, particularly young men, around changing gender norms and sexual behaviors that place young women and men at risk of HIV infection.
- Intervention activities will occur for 3 years
Process Objectives CM

- To raise awareness/increase consciousness
- To identify a shared concern
- To develop organizational strengths to address the shared concern
- To develop leadership to address the shared concern
- To promote action around the shared concern
Types of activities

- One on One Engagement
- Outreach
  - Education in places males gather (Shabeens, Soccer Matches, Taxi Ranks, etc.)
- Workshops
- Mobilization
  - Rallies
  - Murals
- IEC—Digital Stories
- Media—Radio? Probably not—contamination
Assessments for CM

- Baseline and 36 months: CAPI
  - Men and women 18-35 years across all 24 villages

- Measures:
  - Gender Equitable Men’s Scale (GEMS)
  - Mobilization (R21 Lippman)
  - Collective efficacy
  - Social cohesion
  - HIV risk behaviors
  - IPV
Allocation of Intervention

- Randomizing villages
- Using propensity scores to ‘match’ on key co-variates associated with the outcome (Xu and Kalbfleisch. Propensity Score Matching in Randomized Clinical Trials 2010 Biometrics 66:813-823)

Contamination/Spill-Over
- Avoid in Design
- Choice: Group more villages together so fewer units but less change of contamination OR More units/villages but greater change of contamination?
  - Cluster trial not good choice ONLY to avoid contamination (Torgerson BMJ 2001)
- Little evidence that contamination is a problem in trials of educational interventions BUT few studies report extent, nature and effects of contamination
  - Particular interventions (Media) have a high likelihood of contamination favor cluster randomization
  - Methods to deal with contamination (Inverse Probability of Treatment Weights—see Robins J, Cole S, Wang Y)
Addressing Contamination

- Ways in design to avoid contamination
  - Mobilization team from the intervention villages
  - Not using radio, TV, newspapers as would reach control villages
  - Avoiding conducting major outreach events in areas where people from multiple villages congregate

- Ways to try and address contamination in the design
  - Making sure we measure baseline exposure to similar programs and Sonke programs
  - Measuring exposure in intervention and control groups at end
Other evaluation issues

- Measuring breadth (range of programs participated in) and depth (hours spent)
- Tipping point—what proportion of the target population needs to be reached to see the desired change in behavior at the community level—50%? Track this.
- Process measures important—
  - Monitoring of program activities weekly—monthly—feedback to the study team
- Qualitative Data to understand context—how and why the program worked (or did not!)
Key Indicators

- Dose delivered per community/month
  - Calculated from the CM activity forms
- Population reached per community/month
  - Calculated from the CM activity forms
- Fidelity
  - Dose delivered/dose planned per community/mo
  - CM topics mapped to CM objectives—semi-annually?
Conclusion

- RCTs often not possible but careful planning of evaluation simultaneous to the intervention can produce stronger evaluation designs and thus evidence for impact
  - Step Wedge can be a valuable tool
- Complex interventions require many pieces of the puzzle to be fit together to tell a story—TRIANGULATION
- Qualitative Data and Case Studies very valuable in understanding how and why a program worked (or didn’t)
- Careful and detailed measurement of exposure and potential confounders (among those exposed and unexposed) key in the analysis