

Evaluating Stabilization Interventions

Annette N. Brown, 3ie

Cyrus Samii, New York University and Development & Governance Impact Group (DGIG)

with Monika Kulma



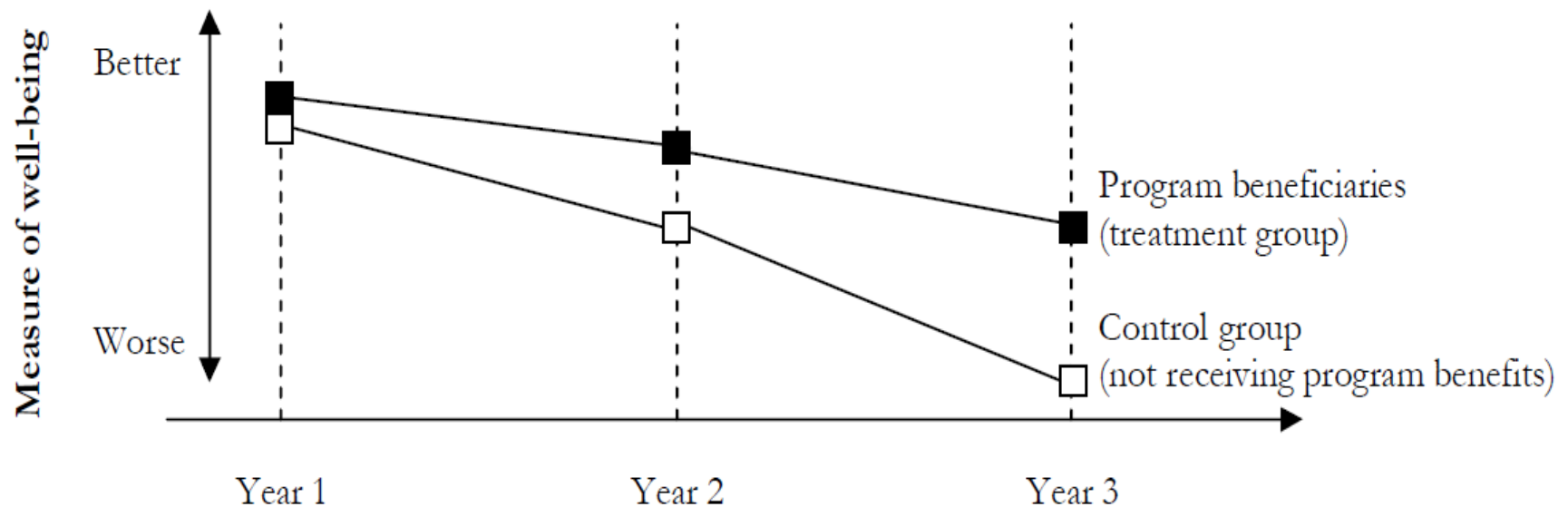
Overview

- Explain motivation for impact evaluation in stabilization program.
- US government experience as an example of current practices.
- Demonstrate what can be done with state of the art stabilization IE examples.
- Focus on identification strategies and measuring outcomes in challenging contexts.

Motivation for IE in stabilization

IE for stabilization requires that usual conditions be met for credible attribution of impacts:

Figure 1: Illustrating the Before-After Fallacy



Motivation for IE for US-funded stabilization programs

Every activity is an opportunity to learn what works, what does not, and why.

2011 USAID Administrator's Stabilization Guidance

USAID Adm. Rajiv Shah



Types of stabilization interventions (US government typology)

- Reintegration
- Civilian police reform
- Community security initiatives
- Peace dividends
- Peace structures
- Peace messaging
- Transitional justice
- Consensus building and dialogue
- Civil society advocacy
- Victims of war

US experience to date

Is the US Producing the Evidence It Needs?

US Stabilization Interventions & Impact Evaluations (IEs) by Category

	Reintegration	Civilian Police Reform	Community Security	Peace Dividends	Peace Structures	Peace Messaging	Transitional Justice	Consensus & Dialogue	Civil Society Advocacy	Victims Of War
US interventions identified	12	60	6	26	8	6	3	12	11	11
US evaluation reports identified	9	3	0	13	4	5	1	6	2	5
US reports meet IE standards	0	0	0	0	0	1	0	0	0	0
IEs for non-US programs	7	0	1	6	1	3	0	2	0	3

- 165 US stabilization interventions identified. **1** IE. ☹️
- Other organizations (World Bank, UNDP, a few I-NGOs) also active in this sector, and some (23) IEs have been produced.

Why is it hard to evaluate stabilization?

- IE requires implementation regularity, but stabilization programs often improvised.
- IE can require considerable planning, but stabilization programs implemented rapidly.
- IE can require careful data collection by implementers, but implementation activities often prioritized over monitoring and data collection for stabilization programs.
- Context makes beneficiary selection sensitive.
- Organizational culture averse to scientific policy making (diplomats, not economists).

The good news

- Despite this, rigorous IE is possible in this area, as the following examples will demonstrate.
- Our goal is to provide examples that will help agencies realize evaluation and learning goals in this sector.

Ex-combatant reintegration in Burundi

(Gilligan, Mvukiyehe, & Samii)

Program

- World Bank/MDRP-sponsored demobilization disarmament, and reintegration (DDR) program after 1993-2004 war.
- Caseload 23,000 in total, including 14,000 ex-rebels.
- Program benefits:
 - 18 months of reinsertion allowances (based on rank);
 - Counseling, including psychological counseling;
 - **“Socio-economic reintegration package”**.
- Intended impacts: economic reintegration that then induces social and political reintegration.

(Gilligan, Mvukiyehe, & Samii)

- Three implementing NGOs. Each NGO assigned a region.
- Africare's implementation was delayed by a year.
- This created a *phased roll-out* scenario, providing a pseudo control group.
- Statistical adjustment to address “incidental” differences across regions.



Ex-combatant reintegration in Burundi

(Gilligan, Mvukiyehe, & Samii)

Outcomes measurement

- Outcomes measured using surveys of ex-combatants.
- Economic reintegration (objective)
 - Income
 - Livelihoods
- Political reintegration (subjective)
 - Preference of civilian life to combatant life
 - Satisfaction with peace accords
 - Support for current government and institutions

Ex-combatant reintegration in Burundi

(Gilligan, Mvukiyehe, & Samii)

Results, with focus on ex-rebels

- Large (20 percentage points) reduction in poverty incidence.
- Moderate increase in attainment of semi-skilled or skilled occupations over unskilled.
- No effect on de-radicalization or political reintegration.

Peace dividends in Aceh

(Barron, Humphreys, Paler, & Weinstein)

Program

- World Bank-sponsored reconstruction and reintegration program after 30 year conflict in Aceh ending in 2005.
- Community-directed development (CDD) mechanisms to allocate resources.
- Intended impacts
 - Enhanced well-being
 - Improved social cohesion
 - Improved trust in government

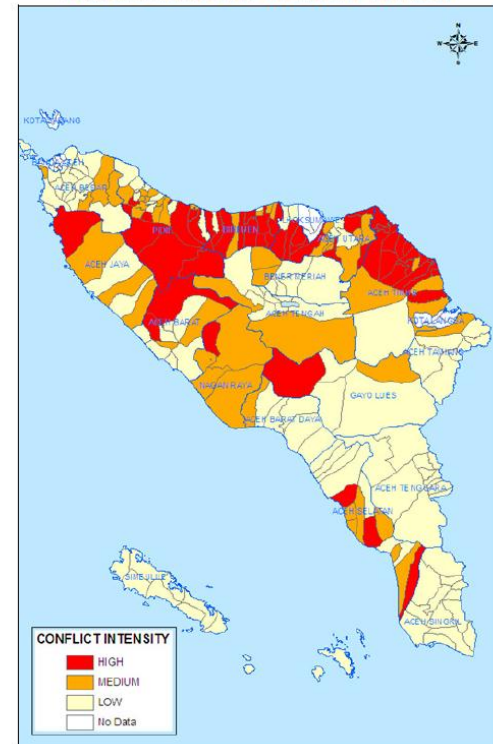
Peace dividends in Aceh

(Barron, Humphreys, Paler, & Weinstein)

Treatment assignment

- Targeted higher conflicted-affected sub-districts in each district
- Conditioned on 60 percent spending criterion

FIGURE 1: CONFLICT INTENSITY IN ACEH



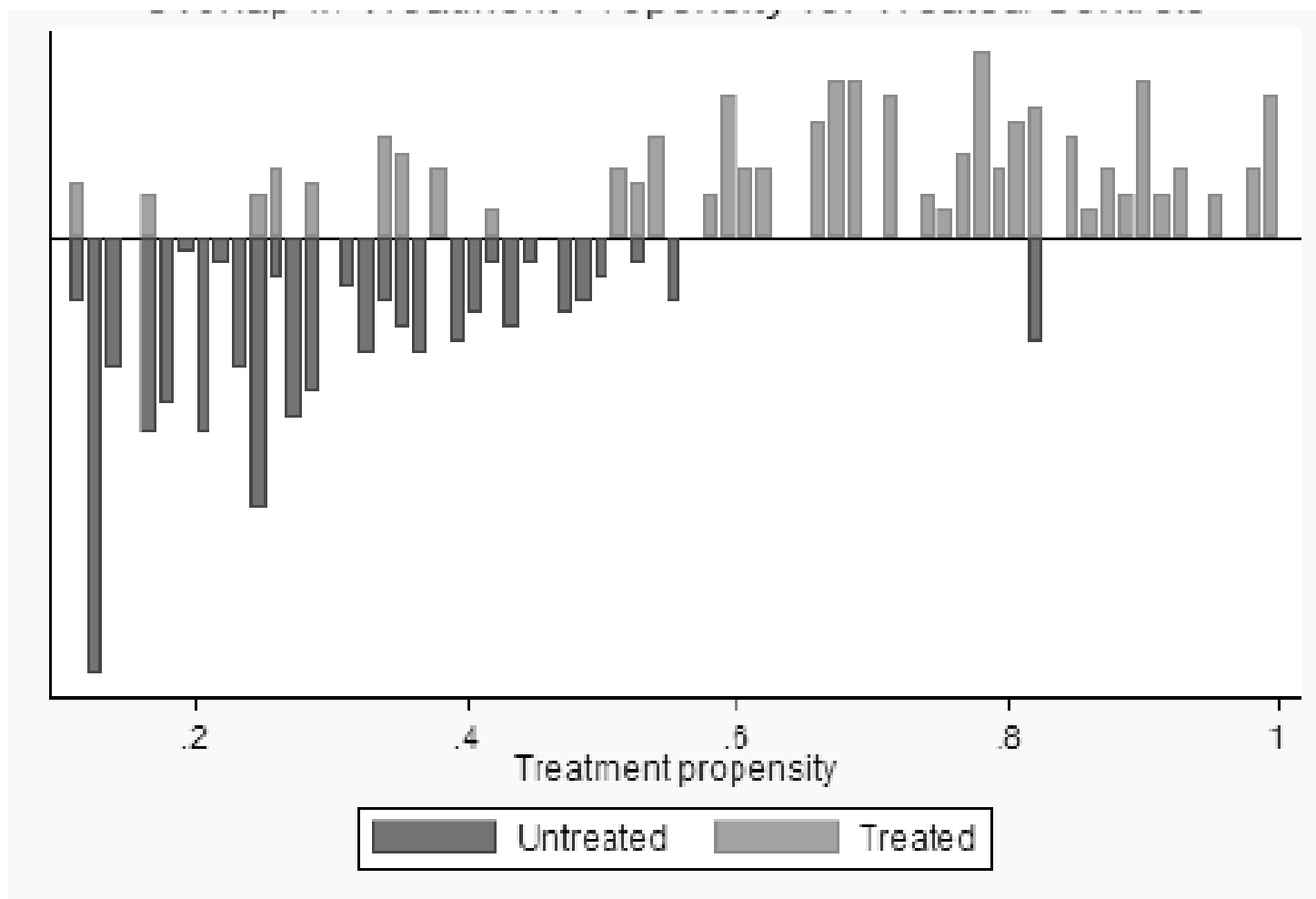
Peace dividends in Aceh

(Barron, Humphreys, Paler, & Weinstein)

Identification strategy

- Propensity score approach to choose a control group
- Use assignment as an instrument—intention to treat
- Regression discontinuity for some of the estimation

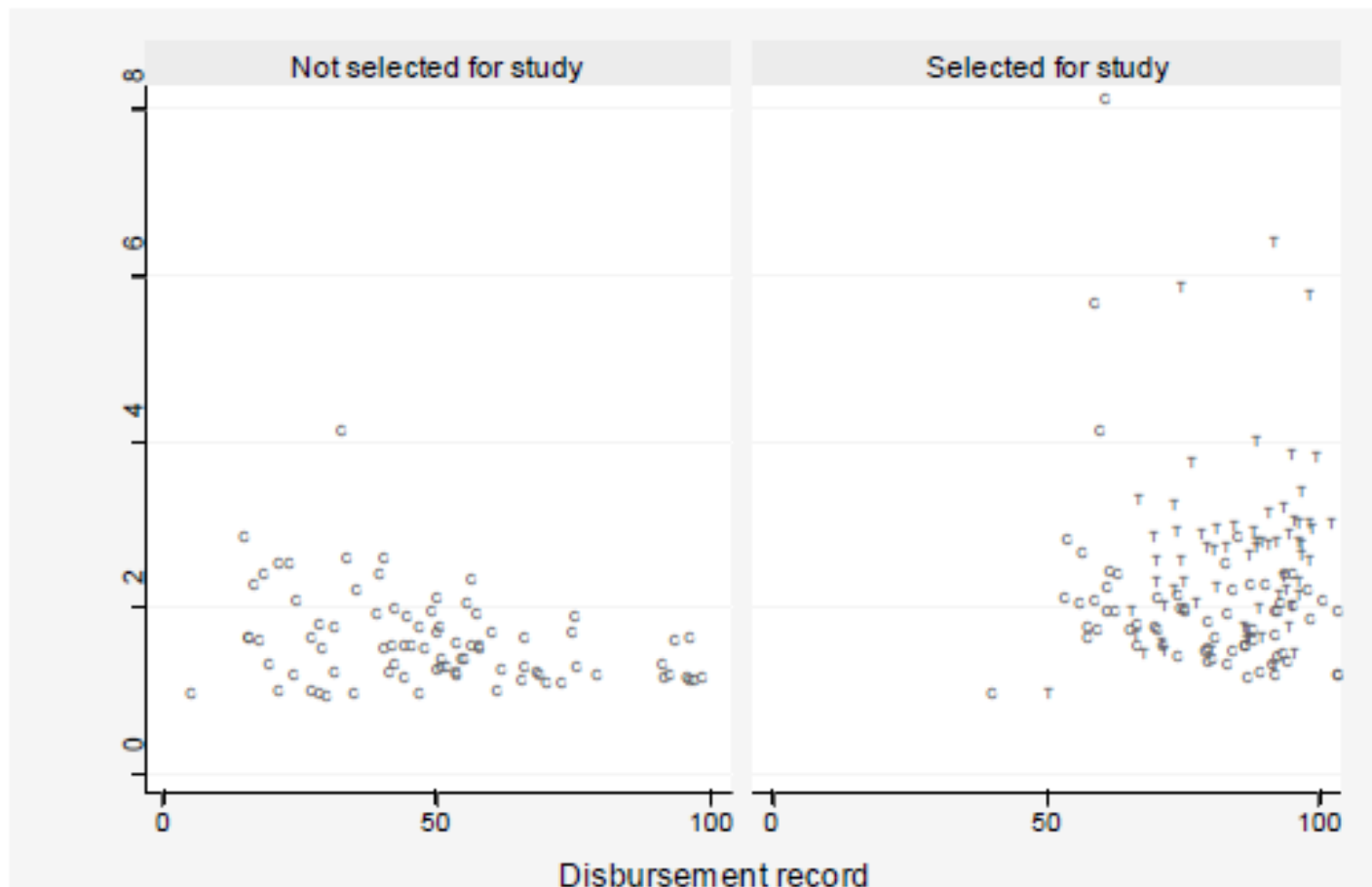
FIGURE 1.2: PROPENSITY SCORES FOR TREATMENT AND COMPARISON UNITS



Source: Authors' calculations

Barron, Humphreys, Paler, and Weinstein. World Bank 2009

FIGURE 1.3: CHARACTERISTICS OF SUB-DISTRICTS SELECTED INTO STUDY



The left panel shows the set of sub-districts not selected into the study and the values of these on two of the assignment criteria. The right panel shows those selected in the study, markers indicate whether sub-districts were treated (T) or not (C).

Source: Authors' calculations

Barron, Humphreys, Paler, and Weinstein. World Bank 2009

Peace dividends in Aceh

(Barron, Humphreys, Paler, & Weinstein)

Outcome measurement using surveys of households and village heads

- Well-being
 - Subjective
 - Poverty rate reported by village heads
 - Subjective perceptions of wellbeing reported by households
 - Objective
 - Assets data reported by households (index)
 - Land use reported by households
 - Wages, employment, education, and health reported by households
 - Public goods reported by village heads

Peace dividends in Aceh

(Barron, Humphreys, Paler, & Weinstein)

Outcome measurement

- Social cohesion
 - Subjective
 - Social distance scale
 - Social tensions
 - Conflict resolution
 - Objective
 - Existence of community projects
 - Participation in associations

Peace dividends in Aceh

(Barron, Humphreys, Paler, & Weinstein)

Outcome measurement

- Trust in government
 - Subjective
 - Satisfaction with village decision making
 - Villagers' role in decision making
 - Confidence in government
 - Objective
 - Contribution game
 - Awareness of government

Peace dividends in Aceh

(Barron, Humphreys, Paler, & Weinstein)

Results

- Numerous positive welfare impacts (11% lower perceived poverty incidence, asset improvements, land use improvements).
- No discernible impact on social cohesion, and even a negative impact on community acceptance of ex-combatants.
- No impact on trust toward government.

Peace messaging in Rwanda

(Paluck)

Program

- NGO (La Benevolencija) reconciliation soap opera called *New Dawn* on radio nationwide.
- Implemented in 2004 to promote inter-ethnic reconciliation after genocide and war.
- Intended impacts
 - Change individuals' own beliefs about out-groups
 - Change perceptions of norms related to prejudicial behavior and ethnic animus
 - Change behavior in the ways encouraged by the program (speak out; cooperate)

Peace messaging in Rwanda

(Paluck)

Identification strategy

- 120 communities were matched into pairs.
- Within pairs, communities were randomly assigned to be treated or control communities.
- In treated communities, listening groups were organized to listen to *New Dawn*.
- In control communities, listening groups were organized to listen to an alternative (health) program during the time that *New Dawn* aired.
- Created a matched-pair randomized control trial using encouragement design.

Peace messaging in Rwanda: outcome measures (Paluck)

- Subjective measures using survey responses to statements about appropriate behavior
- Objective measures using content analysis of focus groups discussions on issues trust
- Objective measures using observation of community negotiation on sharing a radio and cassettes

Peace messaging in Rwanda

(Paluck)

- Results
 - Strong effects on subjects' perceptions of what is socially acceptable behavior (norms).
 - Strong effects on subjects' willingness to dissent in group decision-making.
 - But no impact on subjects' personal beliefs.



Outcome measurement:

General concepts

- Stability is a multi-sectoral phenomenon.
 - Security
 - Political participation and governance
 - Rule of law and justice
 - Economic vitality
 - Social well-being
- Stability operates at the individual, household, community, and national level of analysis.
- Appropriate sectoral focus and level of analysis depends on intervention.

Outcome measurement:

Types of outcomes

- Subjective: attitudes and perceptions
 - Self-reported
 - Grievances: *“are you getting what you deserve?”*
 - Normative beliefs: *“is it okay for your kids to marry members of out-group?”*
 - Hard to observe conditions: *“how much do you worry about theft in the night?”*
- Objective: behavior
 - Self-reported: *“did you vote?”*
 - Observed: *outcome of an activity*

Outcomes and measurement:

Attitudes and perceptions

- Pros:
 - Rather easy to collect;
 - Precise, in principle.
- Cons:
 - Possibly obtrusive and easy for respondents to fake;
 - Susceptible to social desirability bias;
 - Unstable/noisy (people change their minds);
 - Susceptible to “priming”;
 - Sometimes detached from reality;
 - Scales often arbitrary (Likert, Guttman, binary...).

Outcomes and measurement:

Self-reported behavior

- Pros: Easy to collect
- Cons: People have bad memory, possibly obtrusive and easy to fake, susceptible to social desirability bias.

Outcomes and measurement:

Observed behavior

Artificial \leftarrow Artifactual \rightarrow Real-world

- Artificial example: economic games
 - Pros: Incentivizes to act sincerely, measure hard-to-observe traits
 - Cons: Hard to collect, people may not act “naturally” and may misinterpret what they “should” do
- “Artifactual” examples: community resource allocation, collective action task
 - Pros: Realistic, incentivizes to act sincerely
 - Cons: Expensive, much planning required
- Real world examples: satellite data, crime statistics
 - Pros: Direct and unobstrusive, easy to get if systems in place
 - Cons: Expensive if no system in place, hard to assess mechanisms (over-determined)

Lessons for IE design

- Quasi-experimental methods are important.
- Innovations in RCT design can allow RCTs in some cases, as can pilot interventions.
- Statistical methods for dealing with absence of baseline data and small samples important in both experimental and quasi-experimental designs.
- Consistent program implementation and detailed program monitoring data are key.

Lessons for outcome measurement

- Outcome measurement is a key challenge for these evaluations.
- Need a multidisciplinary approach to understand the theories of change and construct outcome measures.
- Different measures of same outcomes allows triangulation—self-reported vs. observed.
- Measures of different outcomes allows testing of theories of change.
- Challenge going forward will be synthesizing the evidence given non-standardized outcome measures.

Conclusions

- Upshot is that with some creativity, IEs *are* possible in this sector.
- IEs so far often reveal mixed effects, suggesting the need to rethink aspects of programming or theories of change.

Thank you.

- Annette N. Brown, 3ie, abrown@3ieimpact.org
- Cyrus Samii, NYU & DGIG, cds2083@nyu.edu