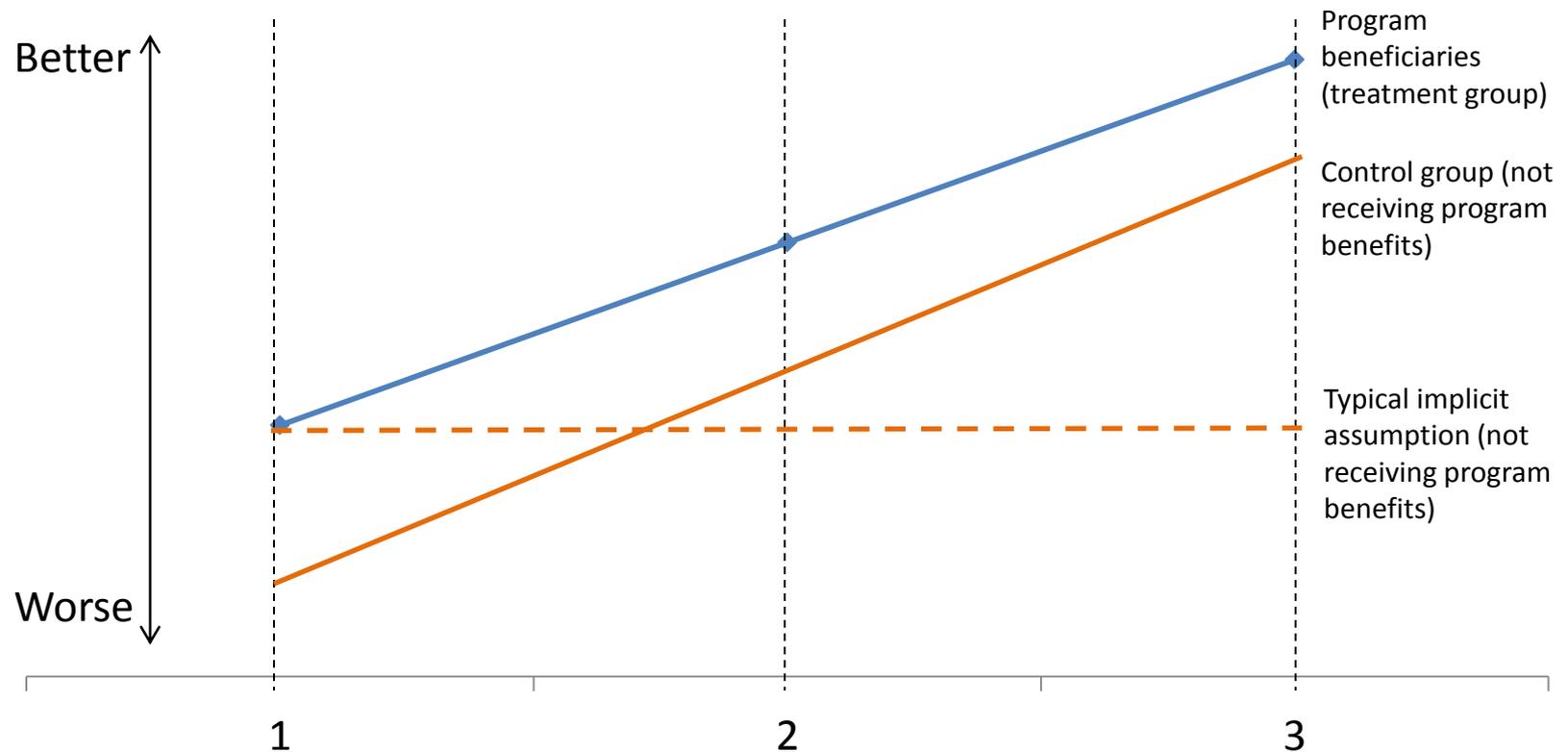# Impact Evaluations in Fragile and Conflict Settings - what can we learn?

Marie M. Gaarder, Manager, IEG Public Sector Evaluation

CIPS / FSRN, University of Ottawa, **January 19, 2015**

IEG WORLD BANK | IFC | MIGA
INDEPENDENT EVALUATION GROUP

Motivation 1: why counterfactual analysis?

An illustration of the Before-After Fallacy

Measure of wellbeing

Better

Worse

Program beneficiaries (treatment group)

Control group (not receiving program benefits)

Typical implicit assumption (not receiving program benefits)

1   2   3

IEG WORLD BANK | IFC | MIGA
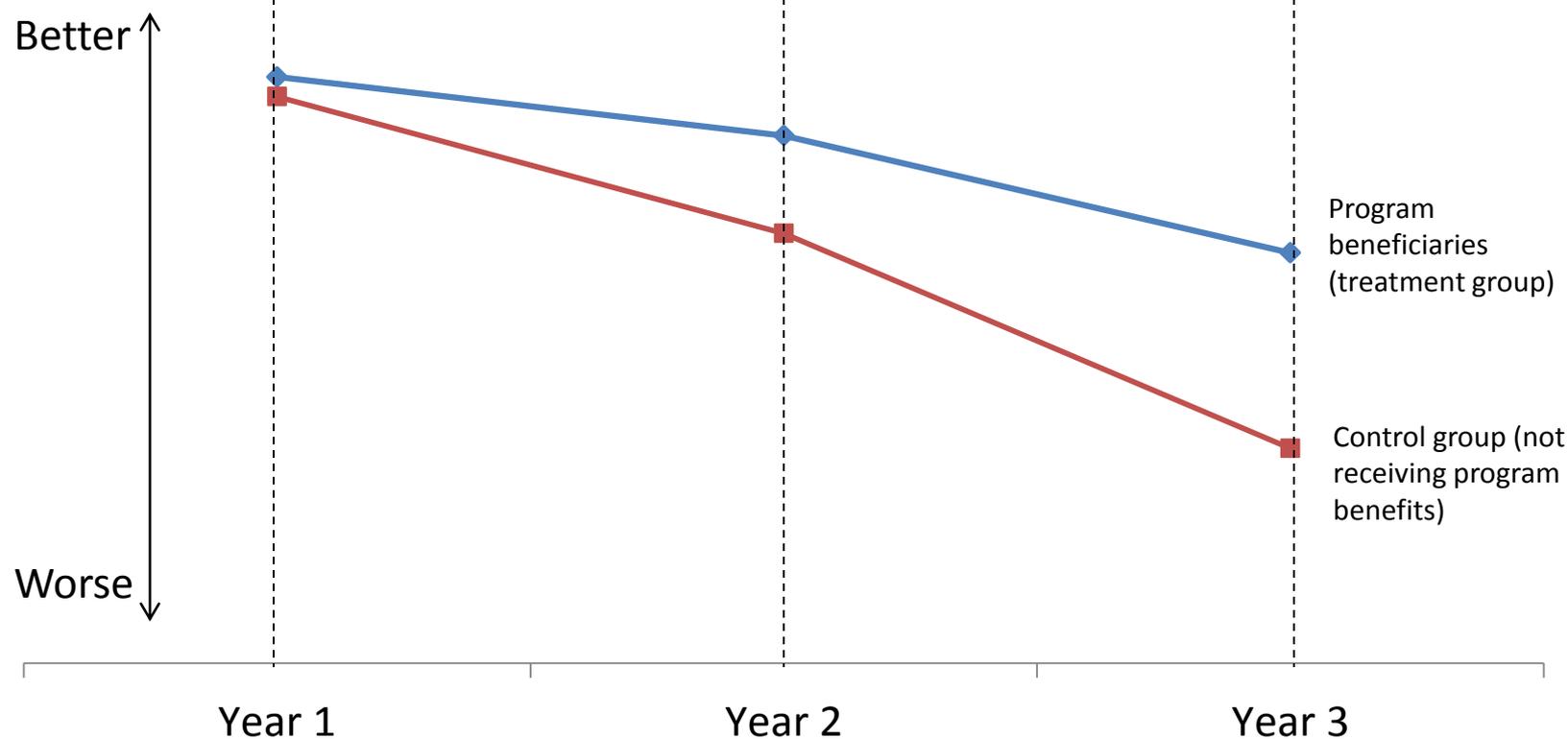INDEPENDENT EVALUATION GROUP

The problem of selection bias

► Program participants are not chosen at random, but selected through
  - Self selection
  - Program placement

► Problem:
  - the beneficiaries of the program are not comparable to the non-beneficiaries
  - They may differ in a way that affects the outcome

Why particularly important in FCS settings?

Stable counterfactual even more unlikely

Measure of wellbeing

Better

Worse

Year 1   Year 2   Year 3

Program beneficiaries (treatment group)

Control group (not receiving program benefits)

Example 1: School-based peace education program

▶ School-based peace education program among Jewish Israeli and Palestinian adolescents was assessed (Biton & Salomon, 2006)

▶ Pre-test and post-test: <u>no change in attitudes</u>

▶ Comparison group that had not received the program: between pre- and post-test <u>mistrust and hatred of Jews had doubled</u>

Example 2: Talk show to reduce conflict

► Democratic Republic of Congo; evaluation of a talk show following a radio soap opera *Kumbuka Kesho* (think of tomorrow) aimed at reducing conflict through community cooperation (Paluck, 2010).

► Testing conflicting psychological theories about the effectiveness of discussion to reduce conflict

► Findings: those listeners who were encouraged to discuss through the additional talk show did indeed **discuss more**, but were also found to become **more intolerant and less likely to aid disliked community members**.

Example 3: Ex-combatant reintegration

► Burundi: Ex-combatant reintegration (Gilligan, Mvukiyehe, and Samii, 2010).

► Testing the hypothesis that economic reintegration will induce social and political reintegration.

► Findings: large **reduction in poverty incidence** (moto-taxi, agricultural assets); moderate increase in attainment of semi-skilled or skilled occupations over unskilled; **no effect on de-radicalization or political reintegration**.

Example 4: Community-Driven Development

► Afghanistan: The National Solidarity Programme (Beath et al. 2013)

► Testing hypotheses: Gender-balanced elected Community Development Councils increases villagers participation in local governance and male acceptance of female participation in public life.

► Findings:

- CDDs had few durable impacts on the identity or affiliation of *de facto* village leaders or the role of representative bodies in local governance.

- The mandating of female participation by NSP – and the consequent female participation in project implementation – results in increased male acceptance of female participation in public life (but not female acceptance) and improvements in women's lives.

Motivation 3: IE improves project design

► Impact Evaluation Helps Deliver Development Projects (Legovini, Di Maro, Piza, 2015)

► Delivery of projects with impact evaluation is <u>significantly timelier</u>: common delays are avoided and the gap between planned and actual disbursements is reduced by half.

► Plausible explanations:

- Evidence-based mid-course corrections
- Clearer implementation road map
- Strengthened capacity on the ground, and
- Observer effects

Establishing the counterfactual in conflict-affected settings

► Individual randomization
  • Reintegration program for ex-combatants (Liberia): demand exceeded supply – registrants admitted by lottery

► Group-based randomization
  • CDD-type program (Liberia and Sierra Leone) assigned 42 communities to receive the program of 83 eligible

► Natural experiments
  • Peace workshops for youth (Sri Lanka) used youth nominated but prevented from participating due to unanticipated budget cut as natural control group

► Quasi-experimental
  • Reintegration program for ex-combatants (Burundi) used disruption in roll out of a program to construct a counterfactual – controlled for systematic differences across the areas (matched on individual and community characteristics)

► Regression discontinuity design
  • Uses program eligibility criteria to establish the counterfactual (e.g. community fragility index used for selection into program – those just below the cut-off for participation are a good comparison group)

IEG WORLD BANK | IFC | MIGA
INDEPENDENT EVALUATION GROUP

Introducing the 'counterfactual mindset' in the development business

► Typically, development agencies' targets are expressed as outcome monitoring indicators, i.e. after – before in intervention/project area

- 15% increase in yield/income/math scores/profits etc

► Was it due to the project? Attribution/contribution?

- **Consider 1:** Is it plausible that the intervention caused the outcome just in terms of the timeline of events?

- **Consider 2:** Can you convincingly argue that nothing else could have affected the outcome? The further to the left in the results chain, so early outcome indicators, the more convincingly this argument can be made (e.g. increase in access to water due to building the only well in the village)

- **Consider 3:** Are time series available so you can document trends in other potentially influencing factors, in project and non-project areas? (e.g. road improvement leads to time savings, despite decrease in fuel prices and increase in traffic volume)?

- **Consider 4:** in your argument about plausible contribution, at least discuss other potential influencing factors even if you don't have data

IEG WORLD BANK | IFC | MIGA
INDEPENDENT EVALUATION GROUP

Outcome measurements for stabilization

► 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?'
- Tolerance: 'it is naïve to have confidence in people belonging to other ethnic groups'
- Perspective taking: 'I try to think of reasons why that group takes a different point of view'
- Hard to observe conditions: 'how much do you worry about theft in the night?'

► Objective: behavior

- Self-reported (did you vote?)
- Self-reported hypothetical response to a situation
- Observed: outcome of an activity (artificial/games; real world data: satellite data, crime statistics, intel on geo-referenced violent incidents)

Adaptation of methods for conflict-affected settings

► <u>Adaptations to the sample</u>: Sudan, Afghanistan (regions lost due to conflict), Liberia (high ranking officers excluded from the evaluation due to potential conflict)

► <u>Timing</u>: Rwandan radio program evaluation – interviews in prisons had to be done earlier due to sudden move to release prisoners

► <u>Question formulation</u>: Sierra Leone CDD – little reluctance to discuss ongoing conflict (and individual roles) but more personal tensions dropped (e.g. marital infidelity)

► <u>Focus group composition</u>: civic education program in Southern Sudan – if social divisions based on sect/ethnic group, single sect (or ethnic group) discussion groups organized

► <u>Enumerators' experience and training</u>: trauma sensitivity, social work background, female enumerators for women in Afghanistan, reconciliation radio program in Rwanda used research assistants with both Hutu and Tutsi backgrounds (message of tolerance)

Ethical and feasibility issues to consider

► Does the evaluation factor in <u>time for delays</u>, the potential for <u>higher attrition</u> (security issues, migration etc)?

► Have interviewers been trained in ethical data collection and conflict sensitive approaches to study participants?

► Does the evaluation team have strong key informants who can provide thoughtful analysis about the security situation?

► Is there a security protocol or guidelines for evaluation staff? Does evaluation staff fall under any organizational protection?

► Have the potential ways that the evaluation may introduce risk and harm to participants, interviewers and implementing partners been adequately considered?

► Is the responsibility of the communication of the findings clarified, is there a communications plan in place and is it conflict sensitive?

High risk, high return?

► Impact evaluation is an expensive way to find out that a program is not being properly implemented (<u>check this first!</u>)

► Particularly useful to carry out on major program hypotheses that have not been verified

► Good impact evaluations build on and use factual information and mixed methods in order to

  - Understand why and how something works or does not
  - develop the theory of change underlying the intervention
  - carry out CBA

► Feed into learning, new programs, systematic reviews

► **"Evaluation Methodologies for Aid in Conflict"** eds: Winckler, O.; Bull, B.; and M. Kennedy-Chouane, London: Routledge, Taylor and Francis Group, Forthcoming 2013