

Overview of Evaluation Methods

Ernest Guevarra

21 January 2020

 ernest@guevarra.io

 ernest.guevarra.io

 [ernestguevarra](https://www.linkedin.com/in/ernestguevarra)

Structure of lecture

- Three case studies (one per hour)

In each case study:

- Examine the various considerations for evaluation
- Discuss the possible methods to use for evaluation and their pros and cons
- Present the actual evaluation design used

Case Study 1

Coverage assessment of direct
nutrition interventions in Liberia

Background

- On 3 February 2014, the Republic of Liberia joined the Scaling Up Nutrition (SUN) Movement
- Liberia had developed a National Nutrition Policy which was designed to complement and strengthen actions set out in the Nutrition Health Policy and the Food Security and Nutrition Strategy
- Liberia's 2012 Poverty Reduction Strategy (PRS) had also identified nutrition as a national priority and an integral element of the overall development agenda

Programme in support of nutrition-specific interventions

- In support of this nutrition policy, funders have invested in a three-year nutrition programme (2017-2019) aimed at tackling child undernutrition. The aim of the programme is to improve coverage of nutrition-specific (direct nutrition interventions) across Liberia.
- Specifically, the programme supports 5 key nutrition interventions: 1) treatment of severe acute malnutrition for children 6-59 months old; 2) vitamin A supplementation for children 6-59 months; 3) promotion of appropriate infant and young child feeding (IYCF) practices among pregnant and lactating women; 4) multiple micronutrient powder supplementation for children 6-23 months; and, 5) iron and folic acid (IFA) supplementation for pregnant women.

What evaluation question/s
would you be interested in
for this programme?

Given the evaluation question/s, what consideration/s should you take into account?

What are the approaches/methods that you would consider in answering the evaluation question/questions?

Actual evaluation question posed by the funder, UNICEF and the Ministry of Health

- Has coverage of the five nutrition-specific interventions increased after the implementation of the three-year nutrition programme in Liberia?

Actual evaluation design

- A before and after coverage assessment in 2 selected implementation areas (Greater Monrovia and Grand Bassa county) in Liberia

Case Study 2

Impact evaluation of World Food Programme's moderate acute malnutrition (MAM) treatment and prevention programme in Sudan

Background

- In Sudan, acute undernutrition is considered one of the most serious but least addressed health problems.
- Out of 213 localities assessed in the 2013 Sudan national nutrition survey, 151 had a prevalence of global acute malnutrition above 10% and 72 localities had a prevalence exceeding the international *emergency* threshold of 15%.
- An estimated 500,000 children (5.3%) suffer from severe acute malnutrition.

World Food Programme's community nutrition integrated programme approach

- Delivering nutrition specific and sensitive services that address immediate, underlying and some basic causes of acute malnutrition
- Within this framework, WFP implements a food-based prevention of acute malnutrition programme in which at-risk children are identified and then provided rations of ready-to-use supplementary food (RUSF) in the form of either a lipid-based nutrient supplement (LNS) or a corn-soya blend (CSB) supplement

What evaluation question/s
would you be interested in
for this programme?

Given the evaluation question/s, what consideration/s should you take into account?

What are the approaches/methods that you would consider in answering the evaluation question/questions?

Actual evaluation question posed by World Food Programme

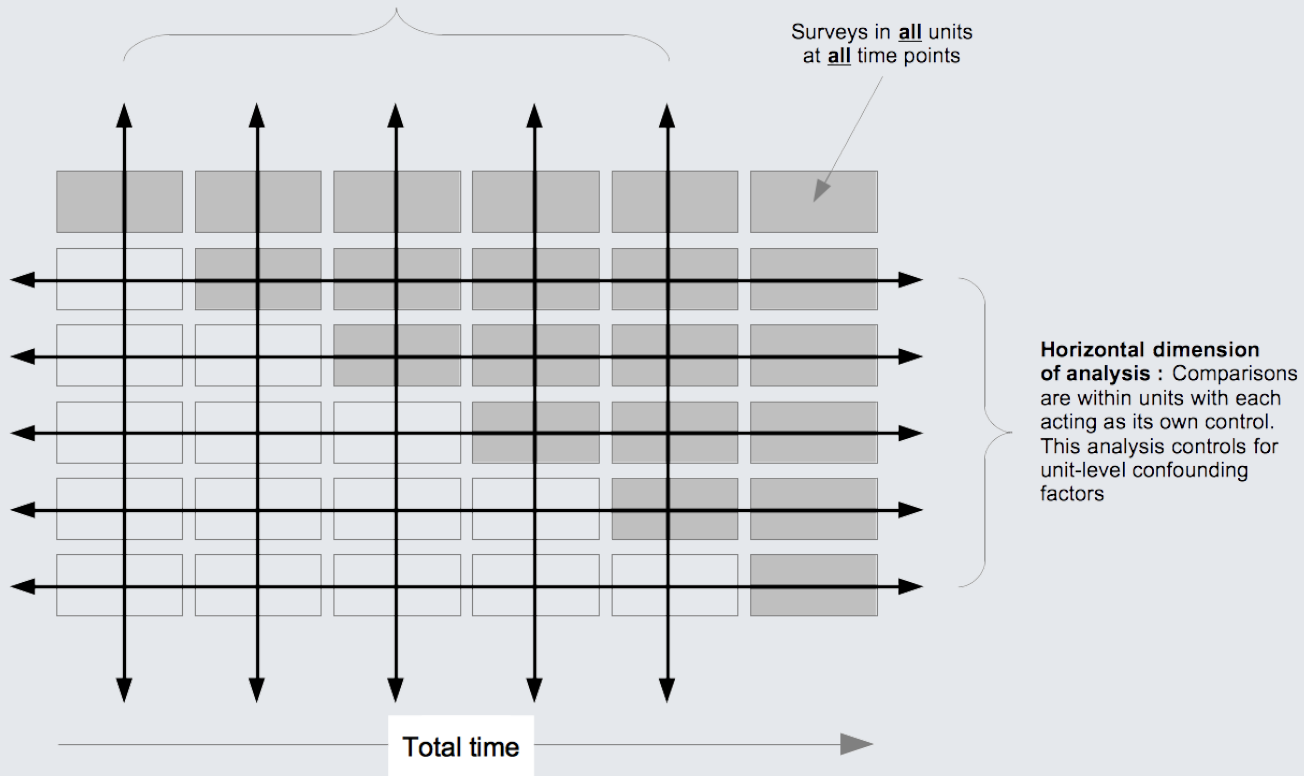
- What is the impact on the incidence and prevalence of MAM and SAM in children under 5 years and pregnant and lactating women (PLW) of different MAM treatment and prevention interventions (i.e. targeted supplementary feeding programme (TSFP) for the treatment of MAM; targeted food-based prevention of MAM (FBMAM), emergency blanket supplementary feeding programme (eBSFP) as a rapid response to crisis for the prevention of MAM, home fortification (HF) for prevention of MAM, and SBCC for prevention of MAM in Sudan?



Actual evaluation design

- A stepped wedge cluster design study (quasi-experimental) to assess impact of programme on prevalence of MAM and SAM.
- A two-arm parallel design cluster controlled study nested into the stepped wedge study to assess GAM incidence.
- A qualitative sub-study to investigate coverage and the effects of social and behaviour change communication (SBCC) in sites across 4 localities selected according to programme status, available data, and accessibility.

Stepped wedge design

Vertical dimension of analysis : Comparisons are between intervention and non-intervention units. This analysis accommodates secular trend and seasonal effects.



 No intervention in this unit  Intervention in this unit

Case Study 3

Assessment of maternal and child cash transfer (MCCT) programme in Kayin and Kayah States, Myanmar

Background

- Kayin and Kayah States remains one of less developed areas of Myanmar and is home to some of the most remote and isolated communities in the country with decades long armed conflicts between Ethnic Armed Organizations and Myanmar Tatmadaw.
- As confirmed by Myanmar Demographic and Health Survey conducted in 2015, children in Kayin and Kayah States are more likely to be malnourished than the average child in Myanmar, with the prevalence of stunting being particularly high.
- Moreover, certain maternal and child health indicators are among the lowest in Myanmar, specifically concerning contraception prevalence, antenatal care visits as well as immunization rates amongst children 12 and 23 months of age.

The Maternal and Child Cash Transfer (MMCT) programme

- Since 2017, the Ministry of Social Welfare, Relief and Resettlement (MSWRR) started rolling out a Maternal and Child Cash Transfer (MCCT) program in Chin, Rakhine, Kayah, and Kayin States through the Department of Social Welfare, in line with the National Social Protection Strategic Plan (NSPSP) and the Multi-Sector National Plan of Action for Nutrition (MS-NPAN).
- The program aims to improve nutritional outcomes for all mothers and children during the first 1000 days of life by ensuring that pregnant women and mothers have improved practices on nutrition, infant, and young child feeding, and health-seeking behaviors.

The Maternal and Child Cash Transfer (MMCT) programme

- The program aims to provide universal coverage to social behavior change communication (SBCC) and a maternal and child cash transfer (MCCT) of 15,000 MMK per month for all pregnant women and mothers with children under 2 years of age.

What evaluation question/s
would you be interested in
for this programme?

Given the evaluation question/s, what consideration/s should you take into account?

What are the approaches/methods that you would consider in answering the evaluation question/questions?

Actual evaluation question/s

- What are the current levels of indicators related to nutrition, infant and young child feeding (IYCF), and health-seeking behaviors of mothers and under 5 children in Kayah and Kayin State?
- What is the impact of the MCCT programme on child stunting in Kayah and Kayin State?
- What is the impact of the MCCT programme on maternal nutritional status in Kayah and Kayin State?

Actual evaluation design

- A cross-sectional survey to assess current levels of indicators related to nutrition, infant and young child feeding (IYCF), and health-seeking behaviors of mothers and under 5 children
- A regression discontinuity (RD) design to assess impact of the MCCT programme on child stunting and on maternal nutritional status

Examples of evaluation methods
or analytical methods used in
evaluation discussed in lecture

Difference-in-difference (DID)

- This is probably the most intuitive of the various evaluation methods discussed.
- In its most simple form, DID is like your typical baseline-endline comparison (as discussed with the coverage assessment plan for the Liberia project) i.e., assess outcome measure before intervention and then measure outcome measure after intervention
- In this comparison, any difference in the outcome measure before and after the intervention can be considered to be partly due to the intervention.

Difference-in-difference (DID)

- With DID, the dimension of having a control group i.e, group not receiving treatment on which the outcome measure is to be assessed at baseline and endline as well, is added to this basic baseline and endline comparison
- The difference in the outcome measure at baseline and endline in the control group can be considered as the change that would have happened regardless of the intervention
- Hence, getting the difference between the difference in baseline and endline in control and intervention groups in theory should give the change that is due to the intervention

Difference-in-difference (DID): Key assumptions

- DID does not account for any other unobserved differences between the control and intervention group that change over time
- DID assumes that the outcome measure changes over time between intervention and control groups i.e., without intervention, intervention and control groups will show equal trends - called parallel trends assumption

Difference-in-difference (DID): Further reading and example

- The Impact of Ethiopia's Productive Safety Net Programme and its Linkages - <https://doi.org/10.1080/00220380902935907>

Methods/approaches for matching

- One of the aims of evaluation methods/design is to be able to describe and obtain a reasonable counter-factual (control) group
- Matching methods as analytic approaches that can be used to be able to match respondents or samples or study participants based on available relevant data.
- Whilst this is not specifically an evaluation method per se, matching methods are used on their own or within specific evaluation designs as a means to specify the most appropriate control group to compare the intervention group with.

Methods/approaches for matching: Types/examples

- Propensity score matching (PSM); An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies - <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3144483/>
- Coarsened exact matching (CEM): Causal Inference without Balance Checking: Coarsened Exact Matching - doi:10.1093/pan/mpr013

Methods/approaches for matching: applications

- Propensity score matching (PSM): The Impact of Ethiopia's Productive Safety Net Programme and Its Linkages - https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1273877
- Coarsened exact matching (CEM): Matching by Coarsened Exact Matching? - <https://views-voices.oxfam.org.uk/2017/06/coarsened-exact-matching/>

Stepped-wedge cluster design

- Commonly used in service delivery type intervention
- Similar in principle to a crossover RCT with the main difference being that the crossover is sequential and progressive over time following the staged or phased roll-out/implementation of an intervention i.e., more clusters are exposed to the intervention towards the end of the study than in its early stages.
- Design is a cluster design hence unit of analysis is not at an individual respondent level i.e., clinic level or hospital level

Stepped-wedge cluster design: further reading

- The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting - <https://doi.org/10.1136/bmj.h391>
- The stepped wedge trial design: a systematic review - <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1636652/>

Stepped-wedge cluster design: examples of application

- From local concern to randomized trial: the Watcombe Housing Project - <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5060144/?report=reader>

Regression discontinuity

- Assigns intervention and control grouping based on cut-off value/criteria for eligibility within a variable that is of a continuous distribution
- Most common example used is in education where this approach is widely used. Eligibility for a special maths class is based on grades on a scale of 0 - 100 where those achieving grades in math of 80 or higher are eligible. We say (assume) that those who just made the cut-off (e.g., those who score between 80-85) are likely to be similar to those who just missed the cut-off (e.g., those who score between 75-79).
- Based on this assumption, these two groups can then be compared either retrospectively or prospectively with regard to the outcome measure of interest.

Regression discontinuity: limitations

- Tends to provide limited external validity as results are only generalizable around the cut-off
- Watch out for possibility of spillover or contamination between control and intervention depending on the treatment (i.e., interventions with rolling recruitment)
- Sample size requirements can be prohibitively high especially if the selected bandwidth of the running variable is very narrow

Regression discontinuity: further reading

- An evaluation of the performance of regression discontinuity design on PROGRESA (World Bank Policy Research Working Paper 3386) - <http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2004/09/09/0000094>

