Common Threats to Impact Evaluation Design

Threat	What Is It?	Why does it Matter?	Example
Selection bias	Selection bias is when participants selected in the study do not represent the population of interest. This could take several forms such as certain groups self-selecting into treatment and/or comparison groups or not being able to take a representative sample of your population of interest. This is an especially important consideration in quasi-experimental studies when an intervention is not randomly assigned. It is critical that the research team consider whether any self-selection is present for either the treatment or comparison group that would impact the results of the intervention.	The implication will depend on the form the selection bias takes. For example, participants self-selecting into your evaluation sample could mean that your findings will not be generalizable to your population of interest, but you still may be able to compare treatment and control groups and generate unbiased findings for your sample only. Self-selection of participants into treatment and/or comparison groups can jeopardize your evaluation because you will no longer be comparing similar groups of participants. This will lead to biased results.	An evaluation is assessing the impact of food vouchers on food security. Villages were randomly assigned to receive the vouchers, and within villages, all households that meet certain vulnerability criteria would be eligible to receive the voucher. In the control villages, households that would meet that vulnerability criteria are identified; however, they do not receive the voucher, since they are the control for the evaluation. In treatment villages, staff do not adhere to the vulnerability criteria and allow additional households to receive the voucher. This jeopardizes the study because in the control villages, all households would meet the vulnerability criteria, but in the treatment villages, they would not, meaning they would be different on average. For example, if the treatment sample now includes households who are more vulnerable than the criteria outlined and thus less food secure, this could underestimate (negatively bias) the true effect of the food voucher.
Attrition	Attrition is when participants drop out of the study so that you are not able to collect data on them. Attrition can take three forms (ordered from least to most severe):	The implications of attrition depend on the type of attrition you have: • **Random attrition:** Sample size is reduced which reduces statistical power.	An evaluation is assessing the effect of an agricultural input distribution program on crop yields. The interview team conducts the survey during the day time when farmers are in their fields. As such, there are no longer farmers in the evaluation sample. The research team will not be

Threat	What Is It?	Why does it Matter?	Example
	• **Random attrition:** when participants drop out but there are no systematic differences between who drops out and who does not. • **Non-random attrition:** when there are differences between the type of people dropping out and staying in, however, there are no differences in who drops between your treatment and comparison groups. • **Differential attrition:** when there are systematic differences in who is dropping out between treatment and comparison groups	• **Non-random attrition:** Sample is no longer representative of the population of interest so findings can no longer be generalized to population of interest. • **Differential attrition:** Treatment and comparison groups are no longer comparable so evaluation findings will be biased.	able to generalize their findings to farmers, only non-farmers. This is an example of non-random attrition. An evaluation is assessing the effect of an agricultural training program on food security. A subset of participants who were less motivated did not complete the training and dropped out of the study. Comparing only those who remain in the treatment group with the control may overestimate (positively bias) the impact of the program. This is an example of differential attrition.
Non-compliance	Non-compliance exists when a participant does not comply with their assigned treatment status. This could mean either those participants that are meant to receive a program, do not take up the program *or* participants meant to be in a comparison group take up the program the treatment group is receiving.	Non-compliance is a threat because it can over or underestimate the impact of your program. When a program improves outcomes, if some treatment participants do not take up the program, this could underestimate the effect. If some comparison participants take up the program, this could also underestimate the effect.	An evaluation is comparing the effectiveness of multi-purpose cash versus multi-purpose cash with supplemental nutrition assistance of food security. The supplemental nutrition is distributed at distribution points. Not all families attend the distributions to receive the supplemental nutrition assistance. This may underestimate (negatively bias) the impact of the multi-purpose cash with supplemental nutrition assistance relative to multi-purpose cash alone.
Contamination	When treatment or comparison groups are systematically affected by an	Systematic influences to either treatment or comparison groups may influence your outcomes of	A matched comparison evaluation is comparing the effect of seed kits relative to seed vouchers on household food consumption. Households in the

Threat	What Is It?	Why does it Matter?	Example
	outside shock, program, or campaign during the period of the study. This is mostly a concern for quasi-experimental studies where treatment and comparison groups are likely not in the same geographic areas and may have more access to outside programming or more susceptible to different shocks.	interest independent of the program. This will introduce bias into our findings through an underestimation or overestimation the effect of the program.	area receiving the vouchers begin to experience an uptick in insecurity such that households are not able to travel to the market. This influences household's ability to buy food, thus reducing their consumption independent of the seed vouchers. Comparing the two groups may underestimate (negatively bias) the impact of the seed voucher program.
Spillovers	People assigned to comparison may benefit indirectly from participants receiving the program in the treatment group. Spillovers may be physical, behavioral, informational, or market wide.	Spillovers influence outcomes of the comparison group that affects our ability to estimate what the true impact of the program would have been.	An evaluation is assessing the impact of a handwashing informational campaign to be introduced into certain villages. Comparison villages that are nearby also learn of the campaign through talking with people from the treatment villages. This subsequently shifts handwashing behavior of the comparison villages. This may lead to underestimation (negatively bias) of the impact of the handwashing campaign.
Behavioral responses to evaluations	Either the treatment or comparison group changes their behavior because of awareness they are partaking in a research study. This could happen for several reasons, including: • Participants receiving program change their behavior due to increased attention. • Participants not receiving program change their behavior	These types of phenomena mean that outcomes of participants will change independent of the program. This will introduce bias into our findings through an underestimation or overestimation of the program effect.	An evaluation is assessing the impact of distribution of sorghum seeds on household food security. Individuals in the comparison group report that their food security status is worse because they expect to be chosen for the next phase of the program. This may lead to overestimation (positively bias) of the impact of the sorghum seed distribution.

Threat	What Is It?	Why does it Matter?	Example
	because they expect to receive access to the program later.		
Violation of parallel trends assumption	This is a form of selection bias that is specific to difference-in-difference designs. The validity of a difference-in-difference design rests on the assumption that in the absence of the program, outcomes of the treatment and comparison groups would evolve at the same rate (even if they begin at different levels). This is why it is strongly encouraged to have multiple rounds of pre-program data to probe on whether this assumption holds.	A difference-in-difference design is not viable unless this assumption holds. In other words, you will have biased results.	A difference-in-difference evaluation is assessing the impact of a farmer training program on household food security. Comparison villages are selected from a nearby region based on similar socio-economic and demographic attributes. Over time, the conditions for cultivating crops are better in the comparison regions relative to the treatment regions leading to improved food security for comparison households. This leads to an underestimation (negative bias) of the farmer training program.

References:

- Threats and Analysis, J-PAL
- Guide 2: Get out the Vote. Why Randomize? J-PAL
- Guide 4: Counseling the Unemployed. Addressing threats to experimental integrity, J-PAL
- Catalogue of Bias, CEBM and Oxford University

This document was made possible by the generous support of the American people through the United States Agency for International Development (USAID). The contents are the responsibility of the Humanitarian Assistance Evidence Cycle (HAEC) and do not necessarily reflect the views of USAID or the U.S. Government.