

## Lecture 12: Primary Data Collection

Friday 07 June 2024

[Meredith Startz](#)

[Presentation Slides](#)

[Recording](#)

**Q: When was this survey conducted in Nigeria? The dynamics have changed, and I am a Nigerian and am curious.**

A: Primary data collection is an interesting exercise. However, researchers have to ensure that they align the data to be collected to the study objectives. You need to think through the variables of interest before getting to the field.

**Q: I was wondering if Prof. Meredith has any recommendations on how to mitigate the risks she mentioned in the data collection. For example, it may be harder to detect effect for certain outcomes.**

A: Do you want to ask her in the Q+A?"

**Q: Out of interest, how did you incentivize participation and accuracy in the Lagos project (for such a large, detailed and personal survey) and how did you avoid that creating bias?**

A: Great question. How about asking her in the Q+A. This is a central issue.

**Q: Hi, great presentation. I have a question regarding collecting primary data using CAWI (Computer Assisted Web Interview) using tools like Microsoft Forms or Google surveys. How do one mitigate the risks of self-selecting bias? Also, if I send survey the entire population but only a sample responded, how do I ensure the sample is representative ex-post?**

A: Do you want to ask Meredith in the Q+A?

**Q: How do you ensure the respondents answer all questions if the questions are many and takes up to 2 hrs?**

A: Yeah, they probably won't! We will ask her some things about that in the Q+A.

**Q: How do you as a young researcher go around determining the sample size when the population you are using keep on changing in the previous literature. For instance in my case I'm working on marine Small-Scale Fisheries (SSF) in Ghana but the population differs in previous works what constitute the population of SSF in Ghana. At least for now, I have gotten 6 or so... Thanks.**

A: This is a huge issue, both enumerator and respondent incentives. There is a very interesting paper by Torsten Figueiredo Walther and Niklas Moneke, showing how child-bearing women's answers are completely distorted by the presence of an extensive module. You find fewer women and lots of women are recorded as aged 14 or 46 in order to avoid filling out the module (which is for ages 15-45, or something like that).