**Mustapha Mai-Iyali Umar**

**Week 1 Community Assignment**

**Questions**

1.a How many strata does the Harmonia survey will have?

*Ten (10) strata. This is because there are 10 provinces with a hundred district each. Therefor each province can be serve as a strata.*

Use the Excel tool provided to answer the next two questions.

1.b List and briefly describe **at least two advantages and disadvantages** of having these many strata compared to having more, for example 100 (one for each district)? (**MANDATORY**)

Advantages

1. Representativeness. Each province is fully represented in the coverage survey. Sufficient data can be generated for analysis of various subgroups. However, if for instance we increase the strata to 100, then we are indeed increasing the precision of the survey results. The higher the number of strata, the higher the degree of data precision. In general, higher number of strata can increase both internal and external validity and reliability of survey data with significant reduction of the margin of error.

2. Manageability – The use of 10 strata has the advantage of keeping the number of strata within acceptable limit for a meaningful study. This can also help lower cost and logistics for the study.

Disadvantages

1. Higher cost. A key issue with having for instance only 10 strata is the possibility of increasing the margin of error. On the flip side, increasing the strata beyond 10 will render the survey unwieldy and increase both financial and logistical cost. We need more resources, people, materials etc

2. Management issues. Increasing the strata will not undoubtedly increase the need for more surveyors and supervisors. Data collation, analysis and interpretation will take longer.

2. How much will the sample size decrease or increase (overall) if you were to design a survey with and with +/-8% precision per strata or +/- 3% precision per strata (as opposed to +/-5% precision per strata used now). Just provide the ‘Total Completed Interviews Needed’. Think about the importance that the desired precision has on overall sample size. (**OPTIONAL**)

* +/-5% precision per strata: 95% (3,550)
* +/-8% precision per strata: 92% (17,600)
* +/-3% precision per strata: 97% (47,243)

3.a What kind of data do you need to collect to complete a table like the one below? (**MANDATORY**)

The planned versus survey implementation data – census figures, survey population, household characteristics information, sample size, total interviews planned versus conducted and responses rates, information by strata, FBR data etc. Essentially this data is collected before and after implementation of the survey. The table represents what was planned and what actually happened.

**Summary Table of Survey Implementation, Vaccination Coverage Survey, Harmonia 2018-2019**

|  |  |
| --- | --- |
| **Sampling frame** | **Questionnaires** |
| YEAR Harmonia Census:  | Household |
| Updated: MONTH/YEAR:  | Children 12-35 months:  |
|  | Facility-based records  |
| Interviewer training: MONTH/YEAR | Fieldwork: from MONTH/YEAR to MONTH/YEAR |
| Households:  | Children aged 12-23 months at time of survey |
| * Sampled:
 | * Eligible:
 |
| * Occupied:
 | * Mother/caregiver interviewed:
 |
| * Interviewed:
 | * Response rate (%)
 |
| * Response rate (%):
 |  |

[**Hint**: see DHS: <https://dhsprogram.com/What-We-Do/Survey-Search.cfm> or MICS surveys: <http://mics.unicef.org/surveys> to have an idea of how a table like this looks with data.]

3.b How does this table relate to potential selection bias and what cautions should you have when interpreting the vaccination indicators if response rate is not 100% and/or several households could not be interviewed? (**MANDATORY**)

This is because it relies on unreliable census data. There is also the problem of children 12 -35 months initially planned for the survey but during the survey there was a difference in the age group assessed. This kind of discrepant situation will undoubtedly raise data quality issues. Consequently it will be necessary to adjust for missing data elements resulting from low response rates or inability to reach the appropriate sample size. As stated in the WHO Coverage Survey Reference Manual, to ensure representativeness of the population, the design weight can be adjusted “to transfer the sampling weight of non-respondents to the survey respondents” (p67). In such circumstance, interpretation of findings must be done with caution highlighting reasons for such caution.

4.a What variables do you need to collect or define in the analysis to estimate a [weighted] vaccination coverage with the first dose of diphtheria-tetanus-pertussis containing vaccine (DTP1) and with the third dose of diphtheria-tetanus-pertussis containing vaccine (DTP3) among children aged 12-23 months in each stratum by: a) ‘documented evidence of vaccination (home-based record – HBR OR facility-based record – FBR)’, b) ‘by recall’, and c) ‘by documented evidence of vaccination + recall’. [Do NOT include the data you need to calculate weights]. (**MANDATORY**)

a) One will required for children with documented evidence of vaccination will include the following variables: age and gender of children in each stratum, card availability, residency types (rural/urban), dates of vaccination for each antigen and maternal education.

b) Variables required for children whose vaccination is based on recall: age and gender, tentative dates of vaccination for each dose, health facility where vaccination was conducted, reason for missing vaccination card, residency type and maternal education.

c) Variables to collect for those with a combination of documented evidence + recall will include: age and gender, dates vaccination given, card availability, health facility, reasons for missing information and maternal education

4.b How does this table relate to potential information bias and what cautions should you have when interpreting the vaccination indicators in surveys with “low” percentage of documented evidence? (**MANDATORY**)

First it is important to compare the expected versus the actual numbers of interviews done and the response rates. The discrepant nature of the target children should also be noted. Statistical analysis should be used to adjust for missing data elements. It is also important to consider reasons for poor HBR + FBR and/or poor recall during interpretation of data. In general terms, one would need to consider multiple factors in the interpretation of the data to avoid drawing wrong conclusions.

 4.c What variables do you need to collect or define in the analysis to estimate a weighted percentage of zero dose children aged 12-24 months in each stratum by documented evidence of vaccination (HBR or FBR), by recall, and by documented evidence of vaccination + recall. (**OPTIONAL**)

Variables of interest for the first group (zero dose children is a stratum with documented evidence) will include: age, gender, information for vaccine for which the child is zero dose, reason(s) for being zero dose, type of residence, availability of vaccination cards, residency (urban/rural) and maternal education.

For the category of children with evidence based on recall only, the variables of interest will be: age, gender, residency (urban/rural), vaccine(s) for which the child is zero dose, reason(s) for being zero dose and maternal education

For the category with documented evidence of vaccination + recall, the variables of interest will be: age, gender, availability of vaccination card, vaccine(s) for which child is zero dose, reason(s) for being zero dose, residency (urban/rural) and maternal education

4.d What would you do with records without information about DTP1 (HBR and FBR did not have a date or tick mark for DTP1 and the mother/caregiver could NOT remember whether the child had ever received a vaccine against diphtheria-tetanus-pertussis and the question was left blank? Explain why. (**OPTIONAL**)

The appropriate option is to assume the child never received the vaccine in question. A valid vaccination should have at least one or a combination of HBR + FBR records + history. However, the complete lack of evidence from the combination of HBR + FBR records and no recall ability, such cases should be considered unvaccinated and be placed in a separate category for separate analysis.

In specific terms, the data should be handled in 2 different ways:

Consider any child with complete missing vaccination records – no HBR + FBR and history of vaccination as unvaccinated and placed in such category for analysis. It would be necessary to collect reason(s) for child not vaccinated and/or poor recall ability from parents/caregivers for further analysis

[**Hint**: you can choose one of these options and explain why: a) not quantify and ignore and exclude the record from both numerator and denominator; b) quantify how many records are like this and convert these missing values to not vaccinated with DTP1; c) not quantify how many records are like this and convert the missing values to not vaccinated with DTP1; d) quantify how many records are like this and convert these missing values to vaccinated with DTP1; or e) not quantify how many records are like this and convert the missing values to vaccinated with DTP1.]

 I will choose to go with option ‘b’. Since there is no FBR and HBR and recall ability is poor, it is safer to consider the child(ren) as unvaccinated. That way one reduces the potentials of over-inflating vaccination records and reducing potential information bias which may affect the credibility of the survey results as a whole.

5.a What information do you need to collect the probability of selection and how does this relate to calculating design weights? (**OPTIONAL**)

1. Comprehensive list population with key variables such as age, gender, residency (urban/rural), level of education 2. Enumeration areas,

5.b What information about the population in each stratum do you need to collect/obtain (from the National Statistics Office, for example) if you want to aggregate the data and calculate national-level indicators? (**OPTIONAL**)

1. Population size

2. Household number

3. Demographic information: age, gender, marital status, household characteristics, composition and size, literacy level, occupation, etc