
Sample Size Calculation

Sample Size

To estimate sample size, you need to know the following parameters:

1. **Estimated Prevalence** for the indicators of interest (GAM, anemia, etc.).
2. **Desired Precision.**
3. **Design effect.** (cluster surveys)
4. **Average household size.**
5. **Proportion of children under 5.** (anthropometry surveys)
6. **Non-response rate.**
7. Size of total population (if the population is small)
8. Level of confidence

Sample Size Calculation [SRS]

To estimate a sample size, the formula is:

$$n = \frac{z^2 \times p \times (1-p)}{d^2}$$

n = sample size

z = constant (1.96)

p = expected prevalence (fraction of 1)

d = relative desired precision (fraction of 1)

Sample Size Calculation [Cluster]

To estimate a sample size, the formula is:

$$n = t^2 \times \frac{(p) \times (1-p)}{d^2} \times \text{DEFF}$$

n = sample size

t = constant (2.045 for df=29 and p=0.05)

p = expected prevalence (fraction of 1)

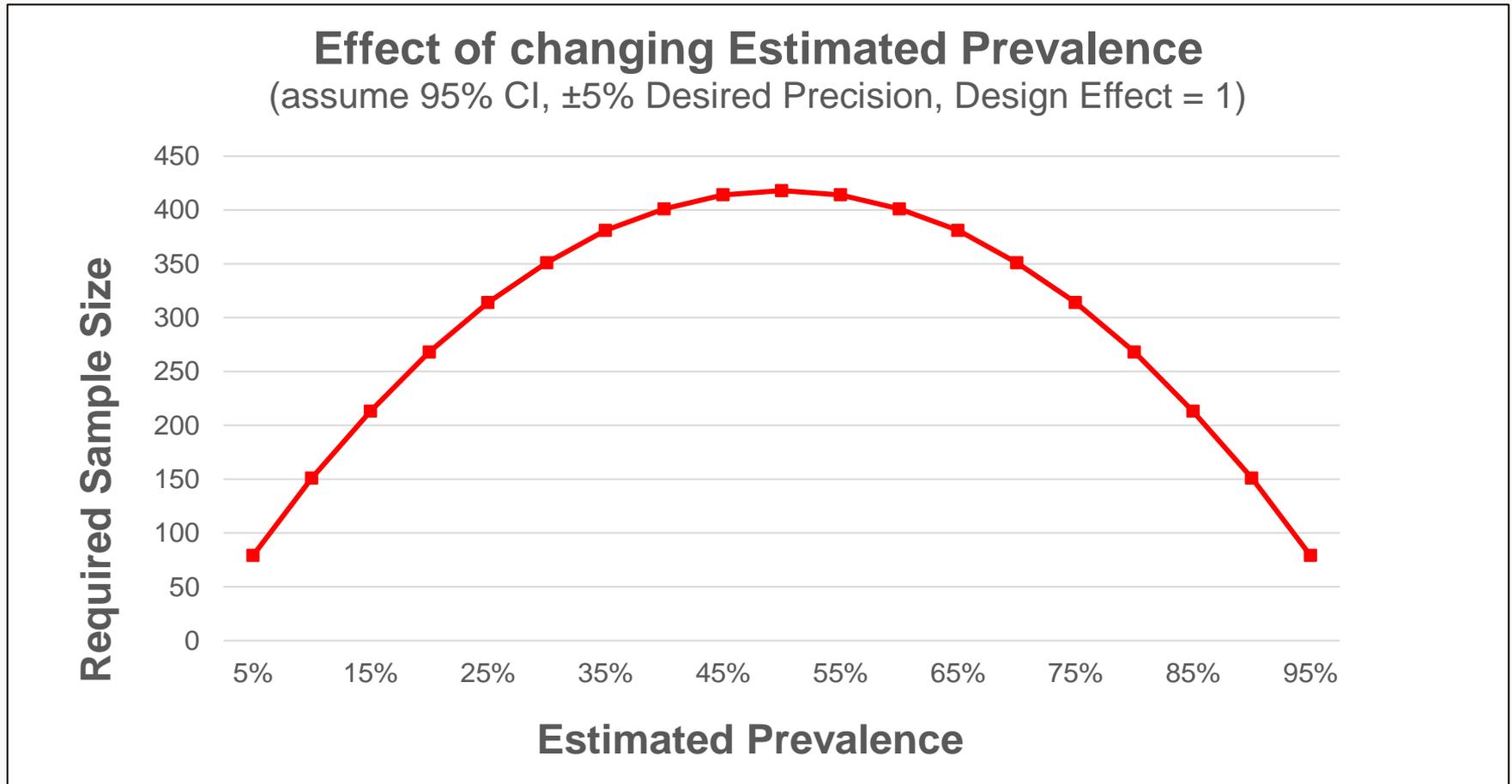
d = relative desired precision (fraction of 1)

DEFF = Design Effect for Cluster Surveys

1. Estimated Prevalence

- **Sources of information:**
 - Previous surveys.
 - Surveillance Data.
 - Qualitative estimates, rapid assessments results.
 - Health workers, religious leaders.
- **Considerations:**
 - Seasonality
 - Aggravating factors
 - Food security
 - Conflict
 - Implementation of nutrition/feeding

1. Estimated Prevalence



1: Prevalence and Sample Size

Estimated Prevalence (%)	As Expressed in Formula [p x (1-p)]	Desired Precision ± %	Sample size (children)
50	0.25	5	384
40	0.24	5	369
25	0.188	5	288
10	0.09	5	138
5	0.048	5	73

Discussion

What should you use if there is no data available on the estimated prevalence of different IYCF indicators in your target population?

2: Desired Precision

To estimate a sample size, the formula is:

$$n = t^2 \times \frac{(p) \times (1-p)}{d^2} \times \text{DEFF}$$

n = sample size

t = constant (2.045 for df=29 and p=0.05)

p = expected prevalence (fraction of 1)

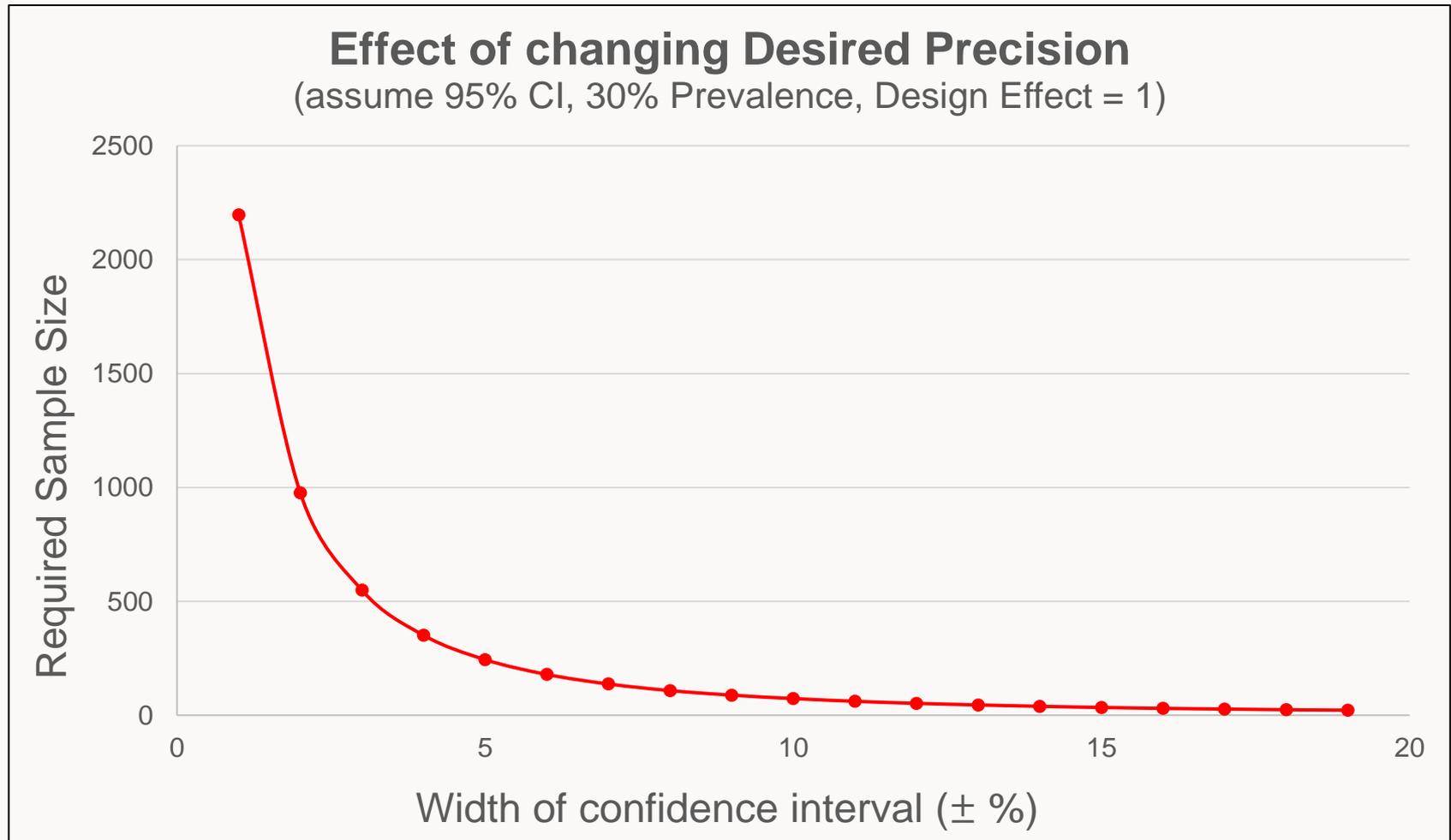
d = relative desired precision (fraction of 1)

DEFF = Design Effect for Cluster Surveys

2: Precision and Sample Size

Malnutrition Prevalence %	Desired Precision ± %	Sample Size (children)
10	2.0	864
10	2.5	553
10	3.0	384
10	3.5	282
10	4.0	216

2: Changing Desired Precision



2: Desired Precision

Precision ↑ (actual number ↓)
Confidence interval narrower →←
Sample Size ↑ ↑ ↑

Unless you have good justification, $\pm 3\%$ should be the **minimum width** of your **confidence interval**.

3: Design Effect

To estimate a sample size, the formula is:

$$n = t^2 \times \frac{(p) \times (1-p)}{d^2} \times \text{DEFF}$$

n = sample size

t = constant (2.045 for df=29 and p=0.05)

p = expected prevalence (fraction of 1)

d = relative desired precision (fraction of 1)

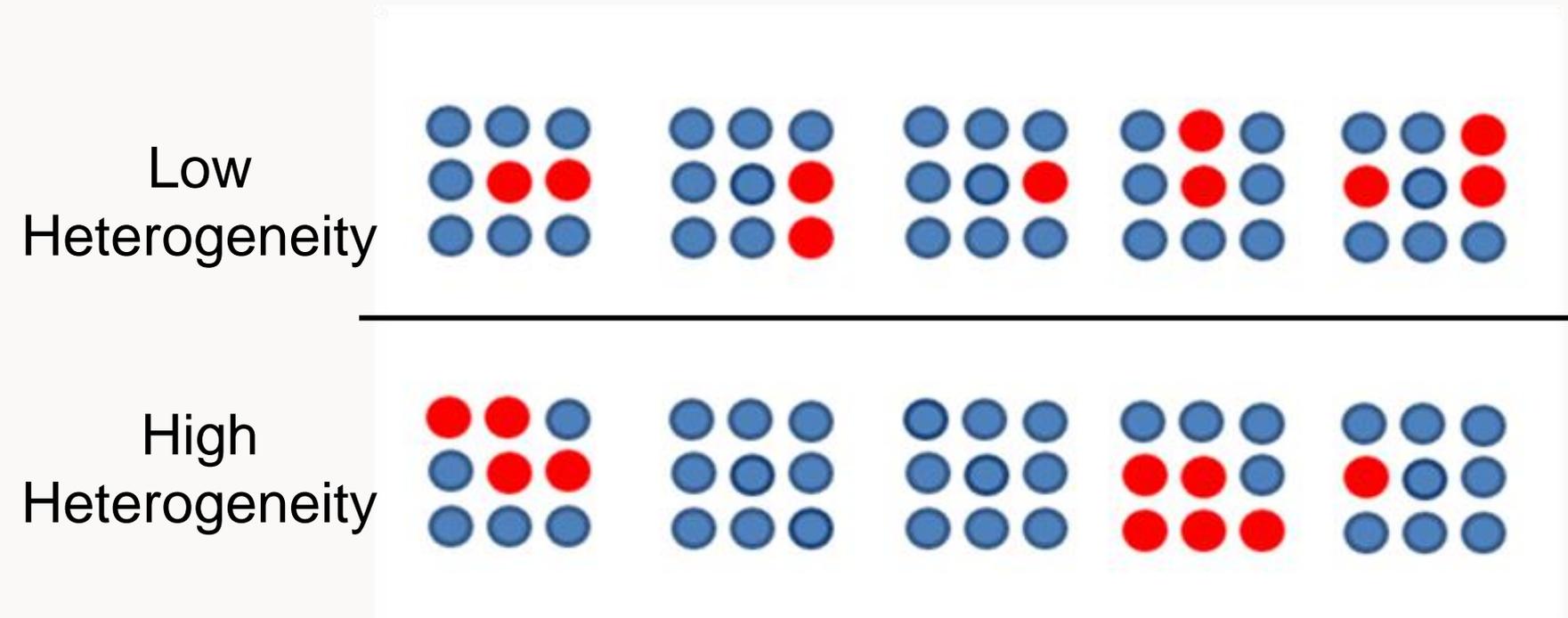
DEFF = Design Effect for Cluster Surveys

3: Design Effect (Cluster Sampling)

- In cluster sampling, it is necessary to **inflate** sample size by a **correction factor** called **Design Effect (DEFF)**.
- Design effect corrects for the effects of **heterogeneity** (differences between clusters).
- In **simple random sampling**, there is no need to correct for heterogeneity; **design effect = 1**.

3: Design Effect (Cluster Sampling)

- Design Effect is a measure of heterogeneity with respect to an indicator of interest



3: Design Effect (Cluster Sampling)

- **Design effect increases ↑ as:**
 - Prevalence in population is more heterogeneous.
 - Cluster size is increased ↑.
 - Prevalence (ex. of GAM) ↑.
- **Selecting a value for DEFF:**
 - Consider the DEFF from previous surveys to help decide which value to use.

Rule of Thumb: DEFF = 1.5

DEFF and different indicators

Different indicators have different magnitude of DEFF.

	DEFF (#clusters = 30)	Indicators examples
Low	<2.5	GAM, stunting, underweight, anemia
Moderate	2.5-6.8	ARI, diarrhea
High	6.8-9.7	Measles immunization
Very high	>9.7	Access to potable water, access to latrines

Converting Sample Size (From Children to Households)

Sample Size

To estimate sample size, you need to know the following parameters:

1. **Estimated Prevalence** for the indicators of interest (GAM, mortality, etc.).
2. **Desired Precision.**
3. **Design effect.** (cluster surveys)
4. **Average household size.**
5. **Proportion of children under 5.** (anthropometry surveys)
6. **Non-response rate.**
7. Size of total population (if the population is small)
8. Level of confidence (always use 95%)

Demographic Data

4. Average HH size.
 5. Proportion of children under 5 years-old (<2 years for IYCF).
- **Sources of data:**
 - ▣ Census.
 - ▣ Past mortality surveys in the same or similar area.

ATTENTION:

Over-estimating the % of U-5/ U-2 and/or the # of persons per HH



May result in having less HHs to visit **BUT** may also result in not having enough HHs to achieve your sample size in # of children.

Converting to # HH

$$n_{HH} = \frac{n_{\text{Children}}}{(\text{HH size} \times \% \text{ of Under 5} \times 0.9)}$$

n_{HH} = sample size in terms of households

n_{children} = sample size in terms of children

HH size = average household size

% of Under 5 = proportion of children under 5 in the population

(x 0.9 since sampling children 6-59 months, not 0-59).

Household Definition

- Context specific
- Agree upon a definition **before** the survey starts.
- Use the definition that is used by the country's government or large scale surveys in the country such as DHS.
- The most frequently used definition is:
"People who slept here last night and ate from the same cooking pot."

6: Non-Response Rate

- Refers to the number of basic sampling units you will **not be able to reach**.
- Refusal.
 - ▣ Sensitivity of questions
 - ▣ Length of questionnaire
 - ▣ Invasiveness of procedures.
 - ▣ Time of day of survey.
- Accessibility.
- Security.
- Absentees.

6: Adjusting for Non-Response

- ▣ Look at previous surveys.
- ▣ May differ by region.
- ▣ May be higher for some measurements (e.g., blood).

6: Adjusting for Non-Response

- ENA uses this estimate to inflate sample size by using the following formula:

$$\text{Final } n = n \text{ of BSU needed } / (1 - \text{NRR})$$

- If non-response predicted to be 10%, divide calculated sample size by 0.9 (100% - 10% = 90%):

$$\frac{\text{Number of HHs needed}}{\text{Expected response}} = \frac{540}{0.9} = 600 \text{ HHs}$$



5 min

Exercise-1

- Calculate the sample size necessary to assess Global Acute Malnutrition among children 6-59 months. (Calculate in number of children AND number of households)
- A previous survey assessing GAM among the same age group revealed a prevalence of 12.0% (9.0-15.0% CI 95%). The study found only small differences in GAM prevalence between the clusters. Since the last survey, responders believe the food security situation has gotten worse.
- Census data indicates that the an average household size in the area is 5.5 The last mortality survey conducted in the same area revealed that 15.5% of the population are under 5 years of age.



5 min

Exercise-2

- Calculate the sample size (both in terms of children and HH) for a IYCF KAP study.
- Based on last survey report, the rate of children ever breastfed was 75% in the area.
- Census data indicates that the an average household size in the area is 5.5 The last mortality survey conducted in the same area revealed that 6.5% of the population are under 2 years of age.



5 min

Exercise-3

- **Calculate the sample size (both in terms of primary and basic sampling unit) for a IYCF KAP midline study.**
- Organization already has the exact list of all the children <2 years in the study location as part of their programme implementation.
- Based on last survey report, the rate of children ever breastfed was 75%, EBF was 40% and MAD was 21% in the area.
- Census data indicates that the an average household size in the area is 5.5 The last mortality survey conducted in the same area revealed that 6.5% of the population are under 2 years of age and 15.5% are children <5 years of age.

Determining number of clusters

Minimum number of clusters

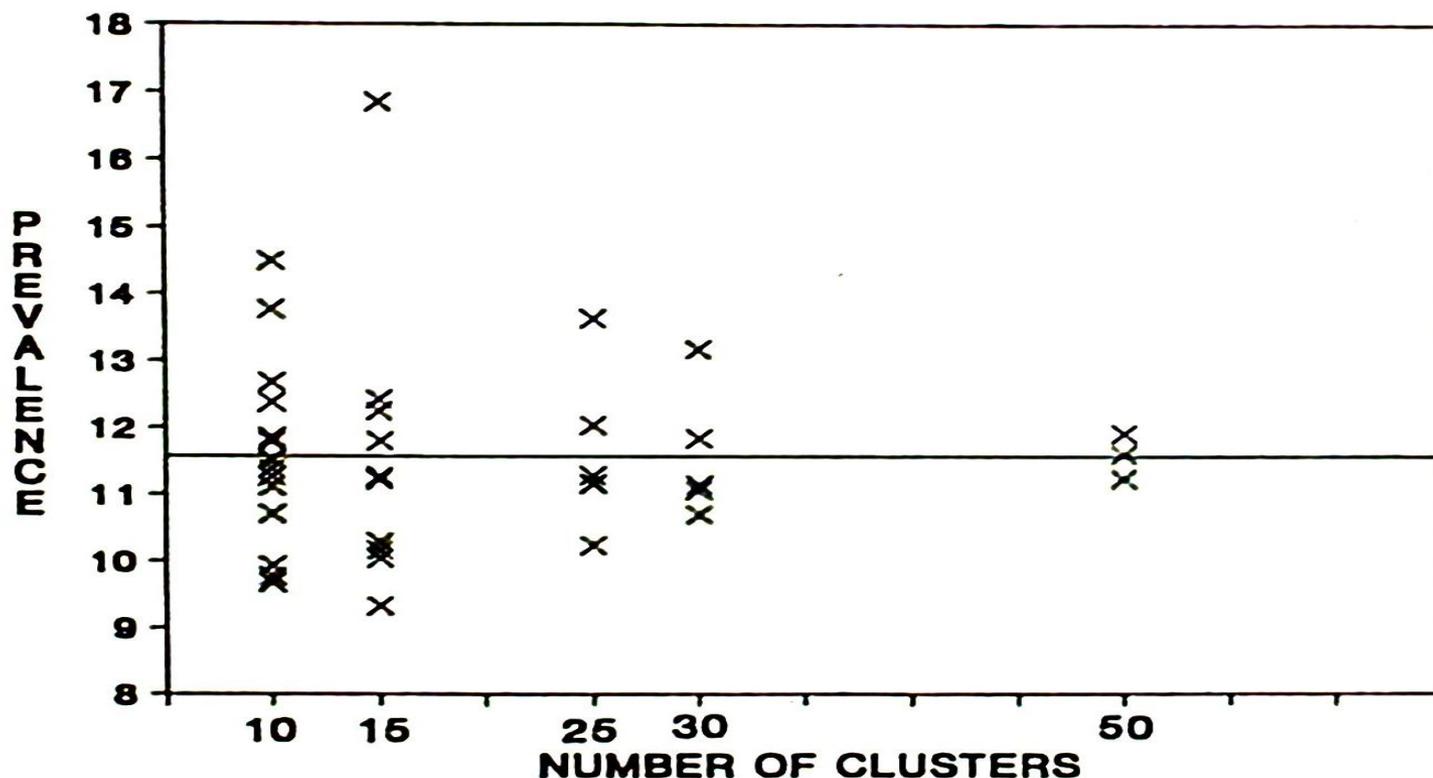


FIGURE 2 Prevalence of low weight-for-height (<80 per cent of median) among children for systematic subsamples of 10, 15, 25, 30, and 50 clusters from a total of 150 clusters, famine assessments in Burkina Faso, Guinea, and Niger, 1984 and 1985. (The solid horizontal line represents the prevalence of low weight-for-age (11.6 per cent) for all 150 clusters.)

Number of clusters

Your survey should ideally include a minimum of 25 clusters

Number of clusters

How many households can you safely visit in a day?

□ Consider:

1. Total length of the work day.
2. Travel time.
3. Time spent on introductions and household selection.
4. Breaks.
5. Average time per survey (per household).

$$\text{Number of clusters} = \frac{\text{Total number of HH in the sample}}{\text{Number of HH to survey per day}}$$

Determining # of clusters

- **Logistics:**
 - ▣ Distance between clusters or from clusters to base.
 - ▣ Whether teams return to the base every night.
 - ▣ Staffing and supervision.
- **Sampling Parameters:**
 - ▣ Expected design effect.



5 min

Exercise

□ **Sample Size= 678 households**

1. Departure from office at **7am** and back at **5pm**.
2. Travel time to reach the village: **1.5h**.
3. Duration for initial introduction and selection of HH: **1h**.
4. Time spent to move from one HH to the next: **3 min**.
5. Average time in the HH: **10 min**.
6. Breaks: **2 breaks of 15 min** each and **1 break of 1h**.
7. Number of teams for the survey: **4 teams**.

How many days will your data collection last, if you want each team to complete one cluster/day?

When to use reserve clusters

- Only use Reserve clusters when:
 - ▣ **10% or more of the original number of clusters were impossible to reach.**

OR

- ▣ Final sample size is less than 80% of the required number.

All reserve clusters should be surveyed if you have **not reached 10% or more of your clusters.**

Reserve clusters

If surveyors are prevented from accessing previously selected clusters after survey starts,

- ENA automatically chooses additional clusters called **reserve clusters** or **replacement clusters**.

Number of clusters	Number of Reserve clusters
25-29	3
30-39	4
40-49	5

Example

- If your target is to survey 40 clusters, but:
 - ▣ You could only access 38 clusters (5% missing), you **do not use the reserve clusters**.
 - ▣ You could only access 36 clusters (10% missing), you should **include all replacement clusters** in your survey, even if there are 5 of them.