Malaria Molecular Surveillance Study Design Workshop

Module 2: Sample Size Calculations Based on Margin of Error

Defining terms

Uncertainty   How unsure we are of the true **population** value based on the information in the **sample**

Precision   The opposite of uncertainty. How tightly we can narrow down our estimate

Confidence interval   A range of values within which we claim the true population value to lie

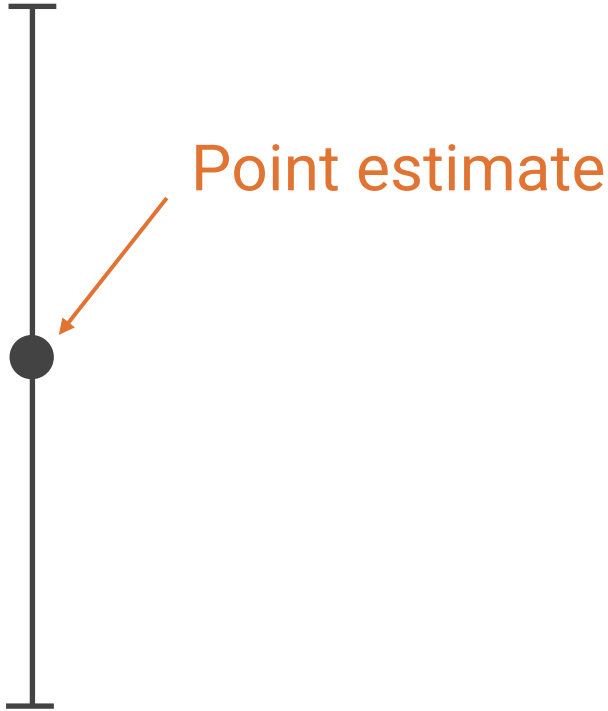**Margin of error**   The distance that our confidence interval extends either side of our point estimate

Low precision

High precision

Low precision
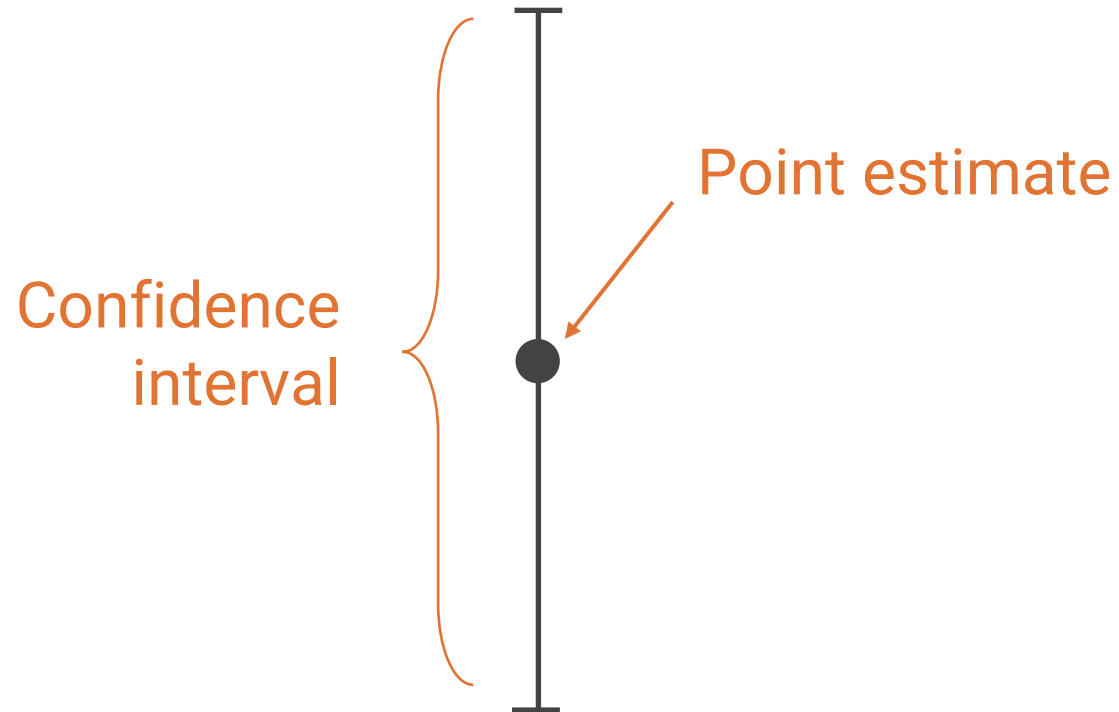
High precision

Point estimate

Defining terms

Low precision

High precision

Point estimate

Confidence interval

Defining terms

Low precision

High precision

Point estimate

Confidence interval
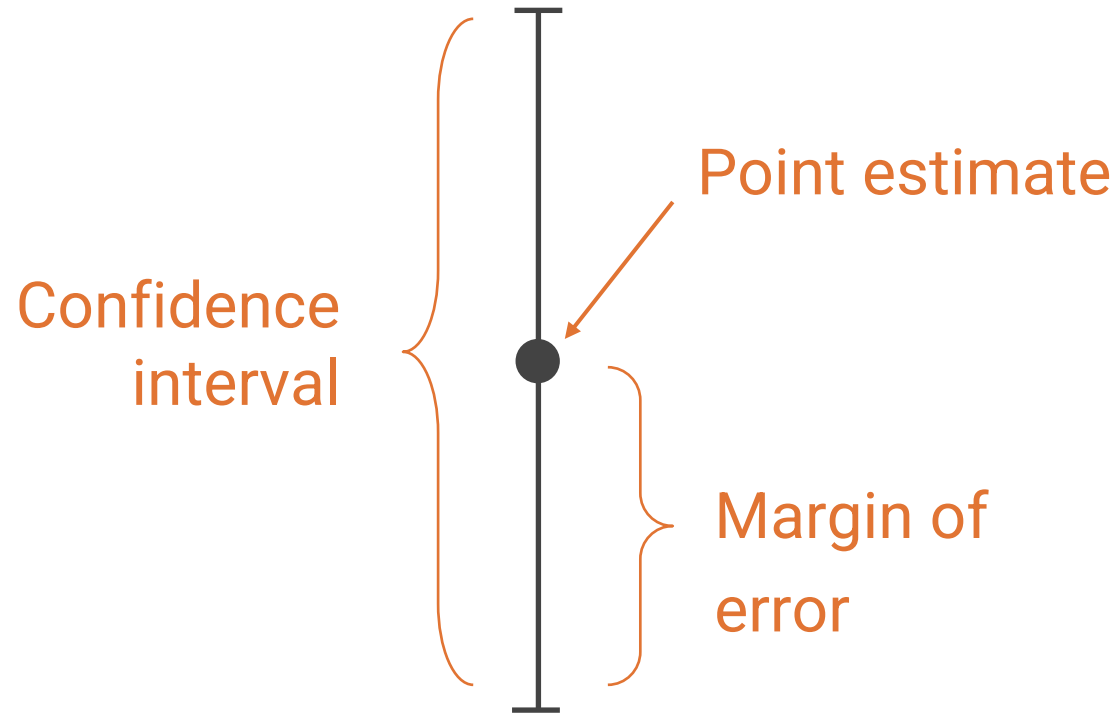
Margin of error

Precision and **accuracy** are not the same thing!
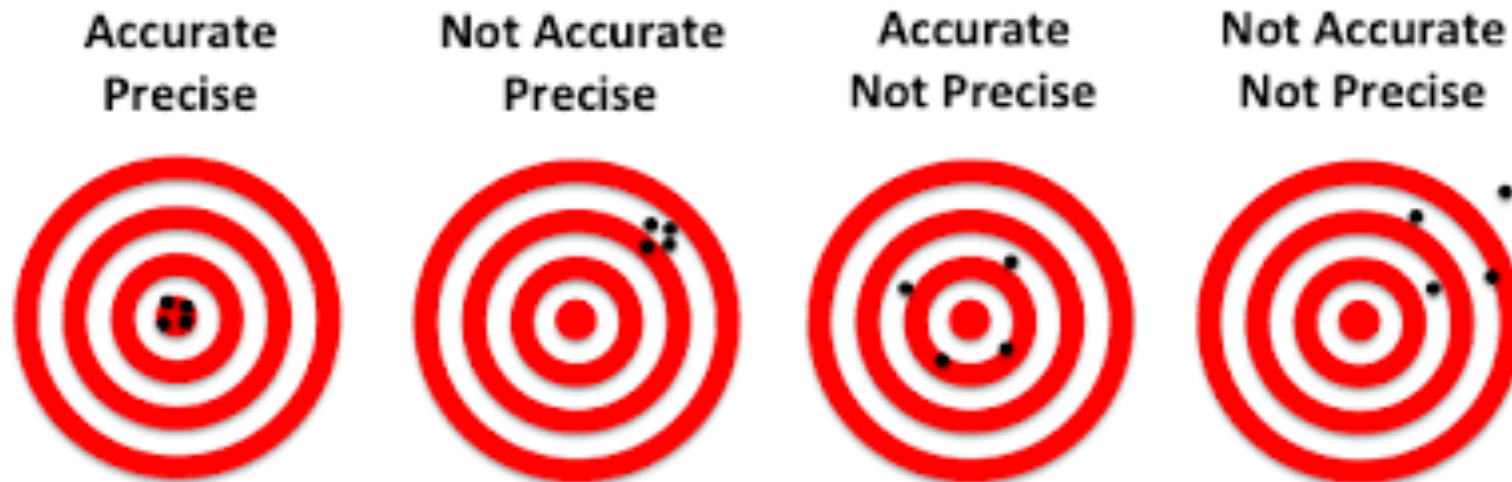
Margin of error

Point estimate

$$\hat{p} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Sample size

$$m = z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Step 1: replace $\hat{p}$ with p

$$m = z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

p is an **assumed** value of the prevalence

Step 2: square both sides

$$m^2 = z_{1-\alpha/2}^2 \frac{p(1-p)}{n}$$

Step 3: multiply by n and divide by m^2

$$n = z_{1-\alpha/2}^2 \frac{p(1-p)}{m^2}$$

Ways to choose p

– Pilot studies

– Historical studies

– Studies in neighbouring regions

– Expert knowledge (e.g. clinical)

Why do I need to assume a value of p?
Isn't p the thing I want to estimate!?

Ways to choose m

– What level of "resolution" does your question require?

– What actions will be triggered based on the results?

– Feedback from sample size calculation: what can you afford? What is logistically feasible?

Background

You are working for the Rwandan NMCP. As part of continuous monitoring, they want to estimate the prevalence of the *dhps* A581G mutation, known to be associated with high level SP resistance. Three years ago, the prevalence was estimated at 15%.

Prevalence: Assume $p = 0.2$ based on previous estimate, plus some wiggle-room

Margin of error: Continuous monitoring, not linked to direct policy change. Values of $m = 0.05$ or $m = 0.10$ may be reasonable

$$n = z^2_{1-\alpha/2} \frac{p(1-p)}{m^2}$$

Large margin of error

$p = 0.2, m = 0.10$

$n = 61.47$

round up to

$n = 62$

$$n = z^2_{1-\alpha/2} \frac{p(1-p)}{m^2}$$

Large margin of error

$p = 0.2, m = 0.10$

$n = 61.47$

round up to

$n = 62$

Small margin of error

$p = 0.2, m = 0.05$

$n = 245.86$

round up to

$n = 246$

$$n = z^2_{1-\alpha/2} \frac{p(1-p)}{m^2}$$

Large margin of error

$p = 0.2, m = 0.10$

Small margin of error

$p = 0.2, m = 0.05$

÷ 2 margin of error

$n = 61.47$

×4 sample size

$n = 245.86$

round up to

round up to

$n = 62$

$n = 246$

Assuming a prevalence of $p = 0.2$

Margin of error (%)	Sample size
1	6147
2	1537
3	683
4	385
5	246
10	62
15	28
20	16

Where is the "sweet spot" on this table?

Assuming a margin of error of $m = 0.05$

Assumed prevalence (%)	Sample size
5	73
10	139
20	246
30	323
50	385
70	323
80	246
90	139
95	73

The highest sample size is always at 50% prevalence

1. Larger sample sizes lead to more precise estimates

2. We can tailor sample sizes to achieve a target margin of error

3. We have to make some assumptions about the prevalence in the population before we measure it

Format: Interactive R code, accessed through the web

– Assist the NMCP of DRC with study design

– Analyse data from a pilot study

– Choose sample size for a follow-up study

– Apply buffering

Format: Interactive R code, accessed through the web

https://mrc-ide.github.io/MMS-SD_workshop/

– Assist the NMCP of DRC with study design

– Analyse data from a pilot study

– Choose sample size for a follow-up study

– Apply buffering

Any Questions?