Malaria Molecular Surveillance Study
Design Workshop

Module 6: Designing a study for multiple end-points

What do we mean by multiple end-points?

We are often interested in designing our study to test more than one outcome

In MMS, our sequencing panels often capture genetic information on ***many molecular markers***

E.g. *pfhrp2* gene deletions + drug resistance markers

How do we design and power our study?

Prevalence estimation: drug resistance mutations

We are collecting data on *pfk13* mutations at a single health facility - our target power is **80%**, what sample size do we need if we assume ***p* = 5%**?

Prevalence estimation: drug resistance mutations

We are collecting data on *pfk13* mutations at a single health facility - our target power is **80%**, what sample size do we need if we assume ***p* = 5%**?

Sample size* = **32 samples**

**We can use DRpower* `get_sample_size_presence()`

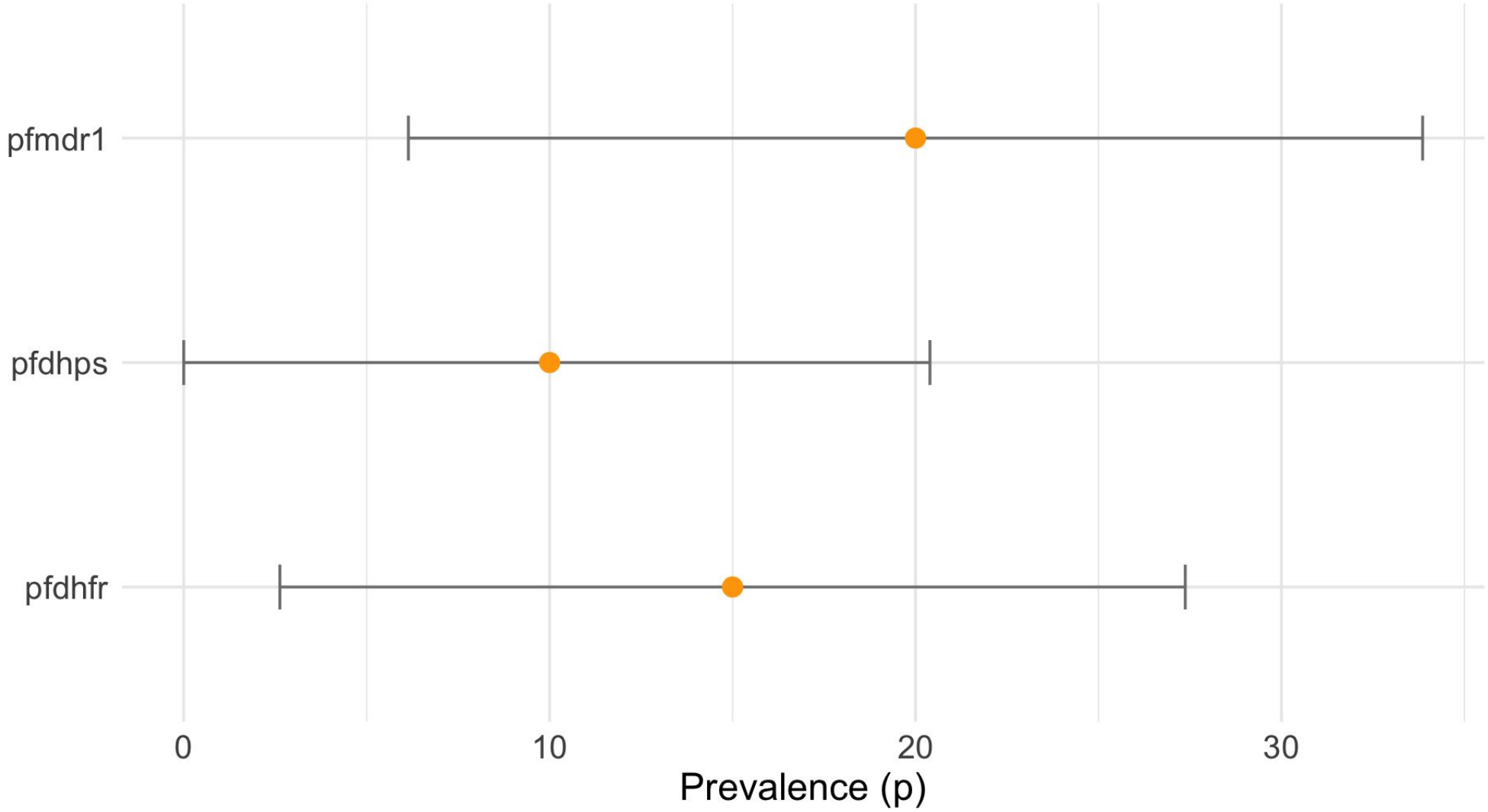
Prevalence estimation: drug resistance mutations

We have now collected data from our 32 samples, but we also want to estimate the prevalence of other drug resistance markers.

Based on expert knowledge, we suspect many of these markers are at high prevalence.

Drug resistance marker	Assumed prevalence (p)
pfdhps	0.10 (10%)
pfdhfr	0.15 (15%)
pfmdr1	0.2 (20%)

High margin of error (CIs)



Designing studies for multiple end-points

1. Specify your primary vs. secondary endpoints ***before*** conducting the study

   Rule of thumb: *If a question is important (primary endpoint) then power for it!*

Designing studies for multiple end-points

1. Specify your primary vs. secondary endpoints ***before*** conducting the study

   Rule of thumb: *If a question is important (primary endpoint) then power for it!*

2. Conduct power analysis and sample size calculation for all primary endpoints. Take the **largest value**.

   This will ensure that you have sufficient power and sample size over all end-points!

Designing studies for multiple end-points

1. Specify your primary vs. secondary endpoints ***before*** conducting the study

   Rule of thumb: *If a question is important (primary endpoint) then power for it!*

2. Conduct power analysis and sample size calculation for all primary endpoints. Take the **largest value**.

   This will ensure that you have sufficient power and sample size over all end-points!

3. **You don't have to be powered to detect everything.**

   It's OK if some things are estimated with low power or precision (eg secondary end-points). They can still act as pilot data to guide future studies.

Going back to our worked example

We design our study to have **two primary end-points**:

1. Detect rare *pfk13* variant mutations (assuming 5% prevalence)
2. Estimate prevalence of *pfdhps* mutation (assuming 10% prevalence)

What steps do we need to take?

Power analysis for both end-points

Calculate minimum sample size required for both end-points (80% power)

- Detection of *pfk13* mutation (assumed prevalence 5%)

  Sample size = **32 samples**

- Estimation of *pfdhps* mutation prevalence (assumed prevalence 10%)

  Sample size* = **139 samples**

  **We can use DRpower* `get_sample_size_margin()`

Which sample size do you choose?

Choose the largest sample size!

We choose 139 as our target sample size (to estimate prevalence of *pfmdr1*) but we are now over-powered for detection of *pfk13* rare mutations

Let's revisit our analysis now….

Choose the largest sample size!

We choose 139 as our target sample size (to estimate prevalence of *pfmdr1*) but we are now over-powered for detection of *pfk13* rare mutations

Let's revisit our analysis now….

We are now powered to detect a prevalence of **2% rather than 5%**
… and 75% powered to detect a **1% prevalence**!

Choose the largest sample size!

We choose 139 as our target sample size (to estimate prevalence of *pfmdr1*) but we are now over-powered for detection of *pfk13* rare mutations

Let's revisit our analysis now….

We are now powered to detect a prevalence of **2% rather than 5%**
… and 75% powered to detect a **1% prevalence**!

There is interplay between the different power analyses for our various primary endpoints. As always, we are looking to strike a balance between what it feasible and what provides the most useful information

Summary

- Studies with multiple end-points are common and it is important to specify our primary end-points *before* conducting our study

- We power for our primary end-points and always select the largest sample size

- We always need to balance power vs feasibility of target sample sizes (eg we don't have to be powered to detect *everything* - pilot data is still useful!)

Activity (1.5 hrs)

Work in groups (group allocation in next slide)

- Scenario-based activity (fictitious!) and ‘role play’
  - Epidemiologist, health facility coordinator, budget officer, statistician (1-2 people)
  - You will be given a specific budget, intra-cluster correlation and ‘fact sheets’ with relevant data

- Together you will design a multi-cluster study for multiple end-points, ***there is no right answer!***

- Last 30 mins - present back to group (1-slide template)

Group allocation

Scenario 1:

1. Reza Niles-Robin
2. Isaac Ssewanyana
3. Dativa Pereus
4. Thomas Katairo

Scenario 2:

1. Agaba Bosco
2. Bernadete Rafael
3. Roland Bamou
4. Lucas Amenga-Etego
5. Mulenga Mwenda

Scenario 3:

1. John Rek
2. Ethan Booth
3. Amadou Niangaly
4. Irene Cevros

Scenario 4:

1. Horace Cox
2. Innocent Ali
3. Billy Ngasala
4. Misago Seth Maze
5. Hinda Doucoure