CORRELATES OF PROTECTION STUDY DESIGN RECOMMENDATIONS

June 5, 2025

START CENTER
STRATEGIC ANALYSIS, RESEARCH & TRAINING CENTER
Department of Global Health | University of Washington

PROJECT TEAM

Helena Manguerra, MPH
PhD Student,
Implementation Science
Project Manager

Sunita Nolan
MPH Student,
Epidemiology
Research Assistant

Cirilus Osongo
MPH Student,
Global Health
Research Assistant

Patricia Pavlinac, PhD, MS
Associate Professor,
Global Health/Epidemiology
Faculty Lead

START OVERVIEW

Leverages leading content expertise from across the University of Washington

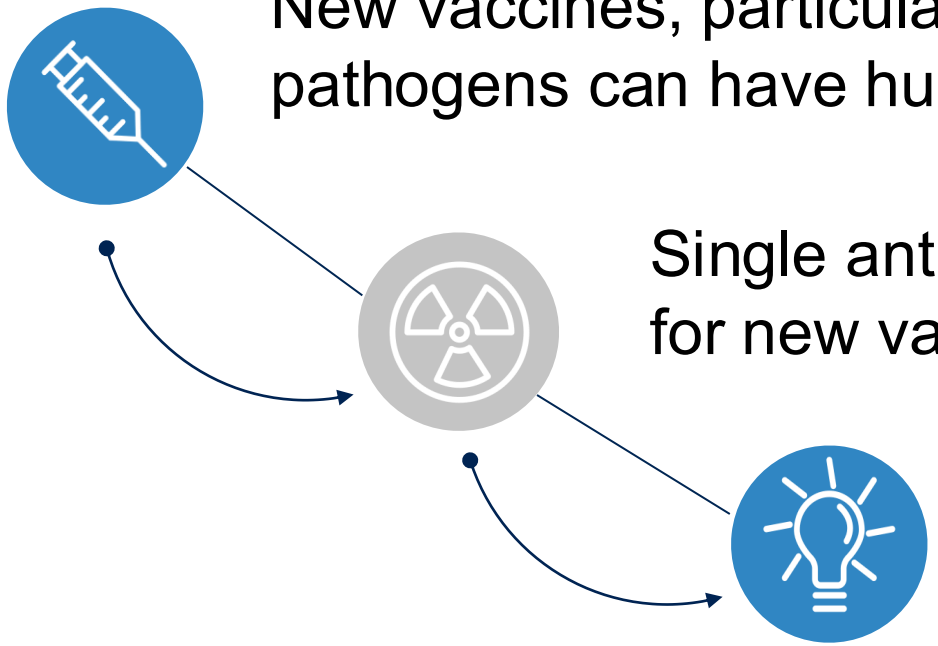
Provides high quality research and analytic support to the Bill & Melinda Gates Foundation and global and public health decision-makers

Provides structured mentorship and training to University of Washington graduate research assistants

AGENDA

1. Project Motivation & Objectives
2. High-level Recommendations
3. Sample Size Analyses
4. Key Considerations for Study Design
5. Limitations & Future Directions

MOTIVATION



New vaccines, particularly combination vaccines, against novel pathogens can have huge impact on reducing health burden

Single antigen phase 3 efficacy studies, which are required for new vaccines, are time and resource-intensive

Well-established correlates of protection provide strong evidence that the vaccine will be protective against the pathogen before running the efficacy studies, and may provide insight into new targets for vaccines

START TEAM OBJECTIVES

Design skeleton of an observational study that could help identify COPs for a wide range of pathogens, utilizing IHME estimates for incidence

Calculate sample size required for the observational study, given assumptions around pathogen burden & COP characteristics

Articulate key considerations that need to be fleshed out in the eventual study

Create calculator with modifiable parameters that can calculate location-specific cohort size

START TEAM OBJECTIVES

WITHIN SCOPE

Design skeleton of an observational study that could help identify COP for a wide range of pathogens across multiple diseases, utilizing IHME data for incidence

Calculate sample size required for the above observational study making assumptions around pathogen burden & COP characteristics

Articulate key considerations for the observational study that need to be fleshed out in the eventual study

Create calculator with modifiable parameters that can calculate location-specific cohort size

OUTSIDE OF SCOPE

Literature review of potential biomarkers for all pathogens

Deep dive into the host immune response for each pathogen to inform optimal timing of blood collection relative to disease and ideal immune markers (i.e. ignore mucosal immunity)

Extensive literature review of pathogen-specific incidence rates beyond what was recommended by experts

Articulation of how to diagnose each disease

Country-specific recommendations for where to conduct the studies

Pathogen- or Biomarker-Specific Literature Review

KEY RECOMMENDATIONS

STUDY DESIGN RECOMMENDATIONS

- 2-Year birth cohort with routine blood sampling and disease surveillance with confirmatory diagnostic testing, including disease serogroup/subtype/sub-serotype
- Stored blood samples (dried blood spots are more feasible)
- Nested case-control studies where cases are disease of interest and matched controls do not have the disease within +/- 2 months of the case occurrence

STUDY SAMPLE SIZE RECOMMENDATIONS

Sample Size: Nested case-control studies

- 40 cases of each disease
- At least 2 controls per case
- If require sufficient power for sub-serotype, then 20 cases of each sub-serotype of interest ae recommended

Sample Size: Birth cohort

- Two separate birth cohorts each powered separately (to ensure one setting not driving the findings):
  - **3,800 children enrolled in South Asia**
  - **2,100 children in Sub-Saharan Africa**

If above is cost-prohibitive, we recommend a birth cohort with 1,900 children from SA and 1,100 from SSA

Likely to identify at least 40 disease cases of

Pertussis
Malaria (SSA only)
Cryptosporidium
ETEC
Shigella
Norovirus
Rotavirus
Campylobacter
Adenovirus
RSV

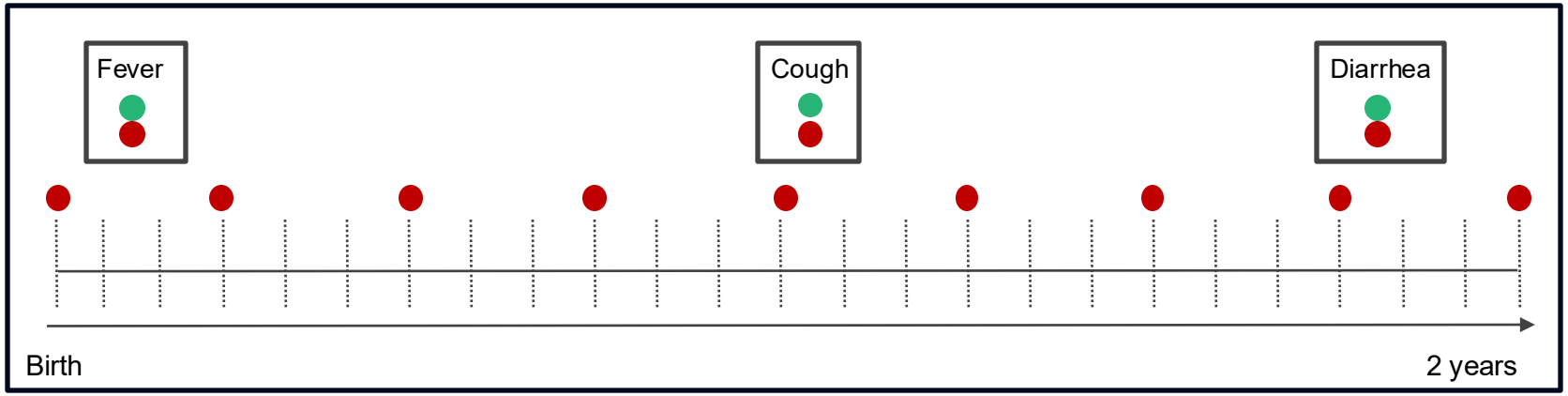
Unlikely to identify at least 40 disease cases of

iNTS
Cholera

STUDY DESIGN OVERVIEW

Birth Cohort (Natural History Study)

2 years of follow-up and active disease surveillance and routine blood collection



Routine study visit at study health facility

Blood sample for antibody testing

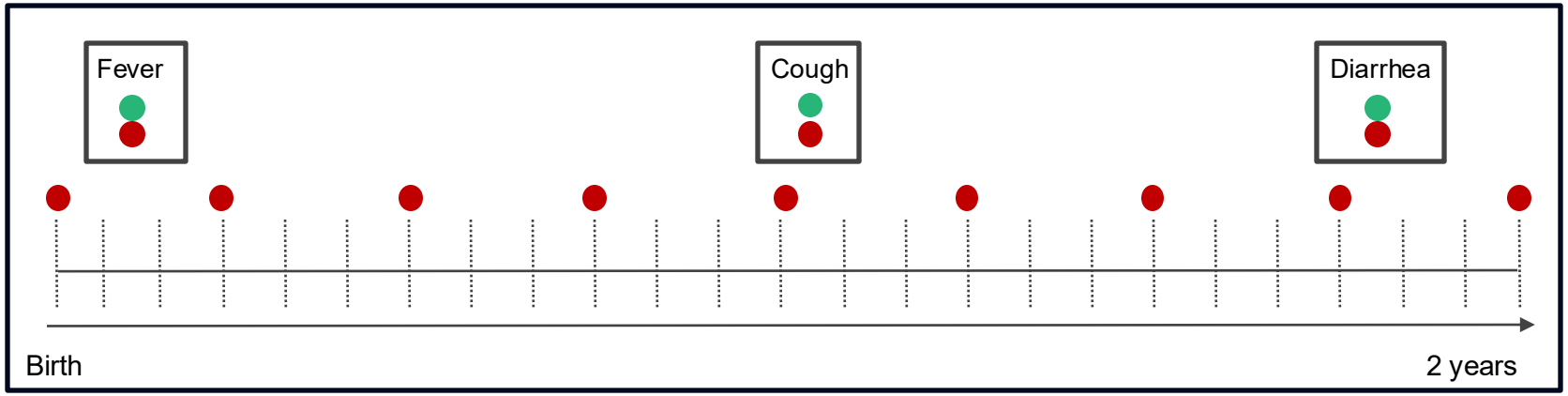
Relevant sample (stool, blood, NP swab) to diagnose disease etiology

Sick study visit

STUDY DESIGN OVERVIEW

Birth Cohort (Natural History Study)

2 years of follow-up and active disease surveillance and routine blood collection

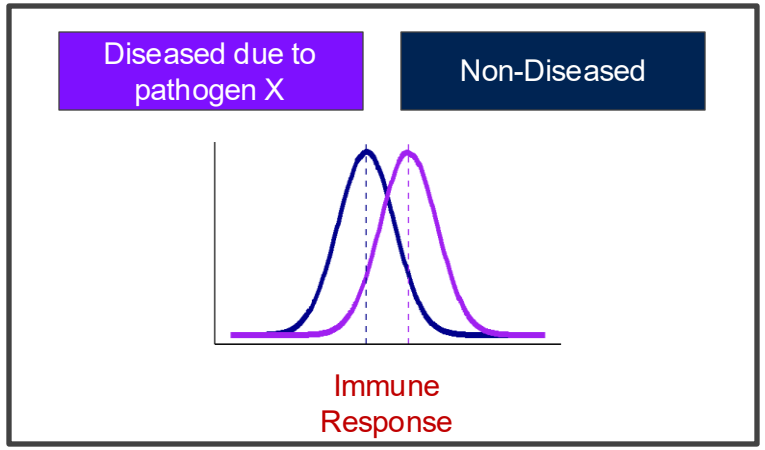


Routine study visit at study health facility

Blood sample for antibody testing

Relevant sample (stool, blood, NP swab) to diagnose disease etiology

Sick study visit



For each pathogen:

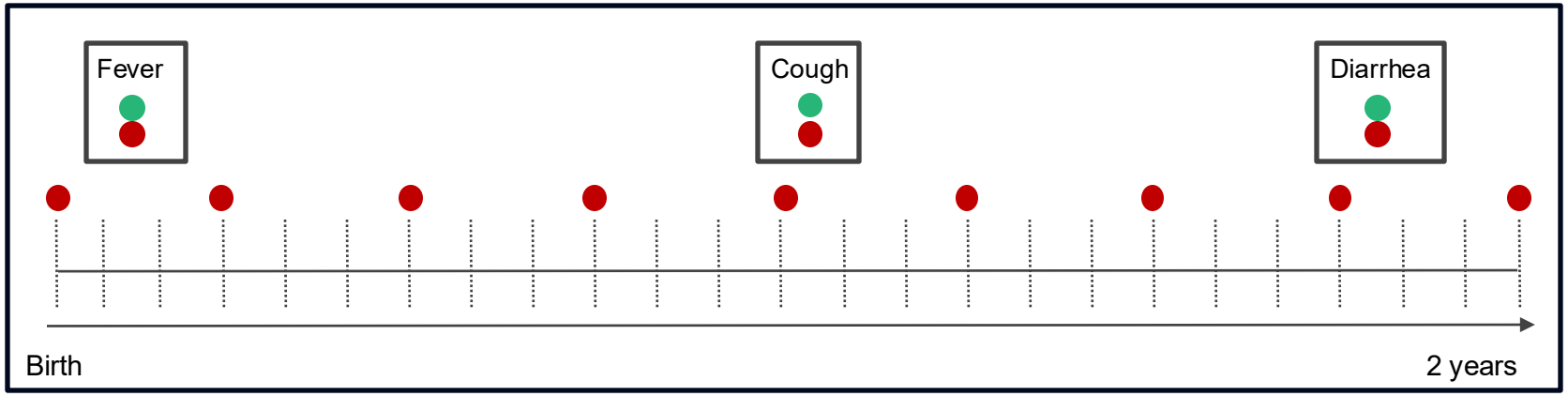
COP case-control analyses

From birth cohort, sample diseased (case) and non-diseased (control) children and compare biomarker levels to identify potential correlates of protection

START CENTER

STUDY DESIGN OVERVIEW

Birth Cohort (Natural History Study)
2 years of follow-up and active disease surveillance and routine blood collection

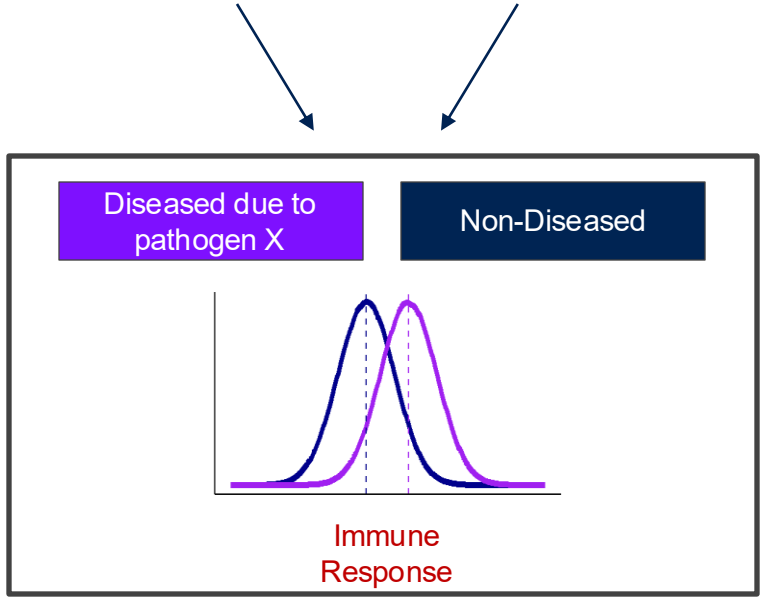


Routine study visit at study health facility

Blood sample for antibody testing

Relevant sample (stool, blood, NP swab) to diagnose disease etiology

Sick study visit



For each pathogen:
COP case-control analyses
From birth cohort, sample diseased (case) and non-diseased (control) children and compare biomarker levels to identify potential correlates of protection

The size of the birth cohort will be determined by the number of children needed to be enrolled to **identify 40 cases of the pathogen with the lowest incidence.**

START CENTER

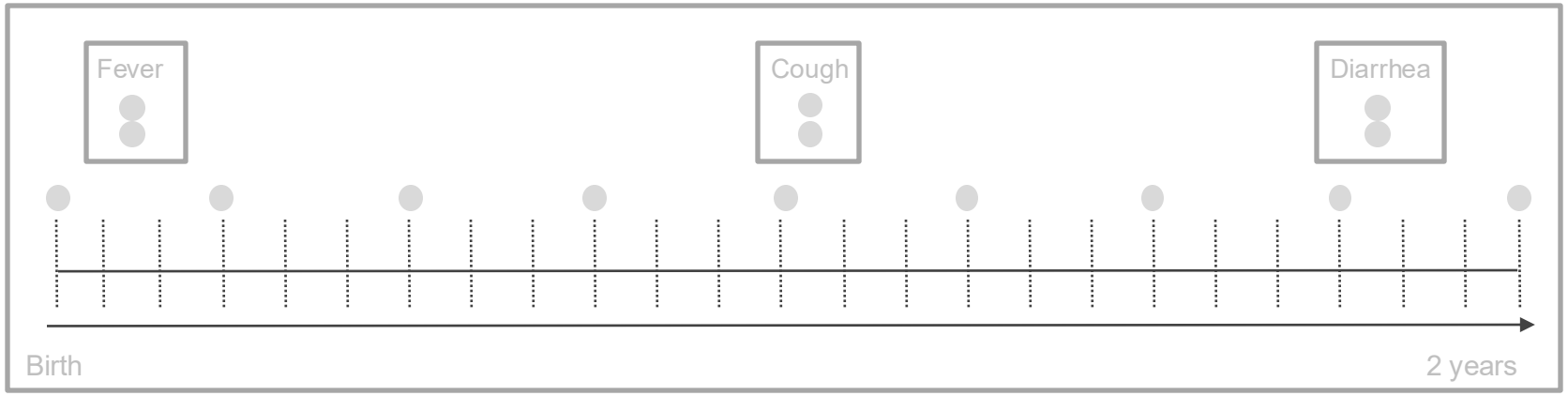
STUDY DESIGN:
COP CASE-CONTROL ANALYSES

STUDY DESIGN RECOMMENDATIONS

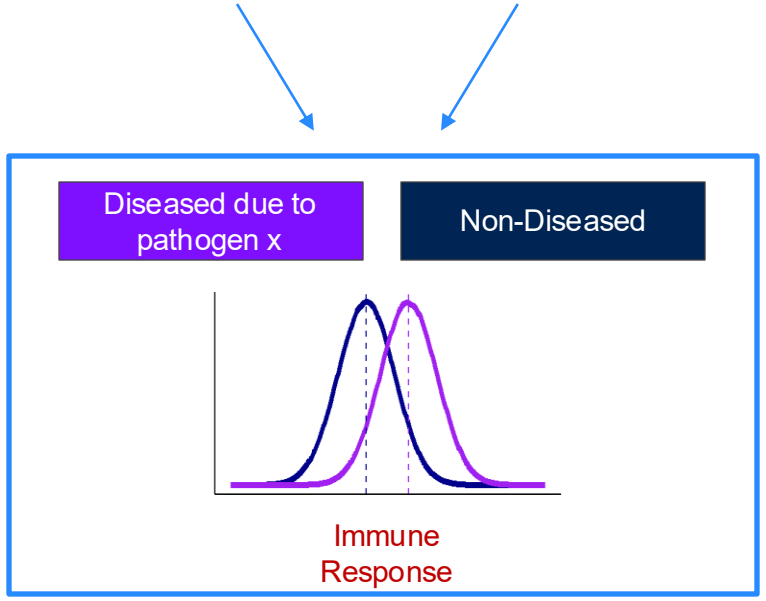
CASE-CONTROL ANALYSES

Birth Cohort

Pool of children enrolled at birth and followed for two years



- Routine study visit at study health facility
- Blood sample for antibody testing
- Relevant sample (stool, blood, NP swab) to diagnose disease etiology
- Sick study visit



For each pathogen:

COP case-control analyses

From birth cohort, sample diseased (case) and non-diseased (control) children and compare blood samples to identify potential correlates of protection

START CENTER

COP CASE-CONTROL ANALYSES

PRIMARY AIM

Estimate odds ratios of the outcome for different levels of an immune marker measured near to and close to the occurrence of the outcome.

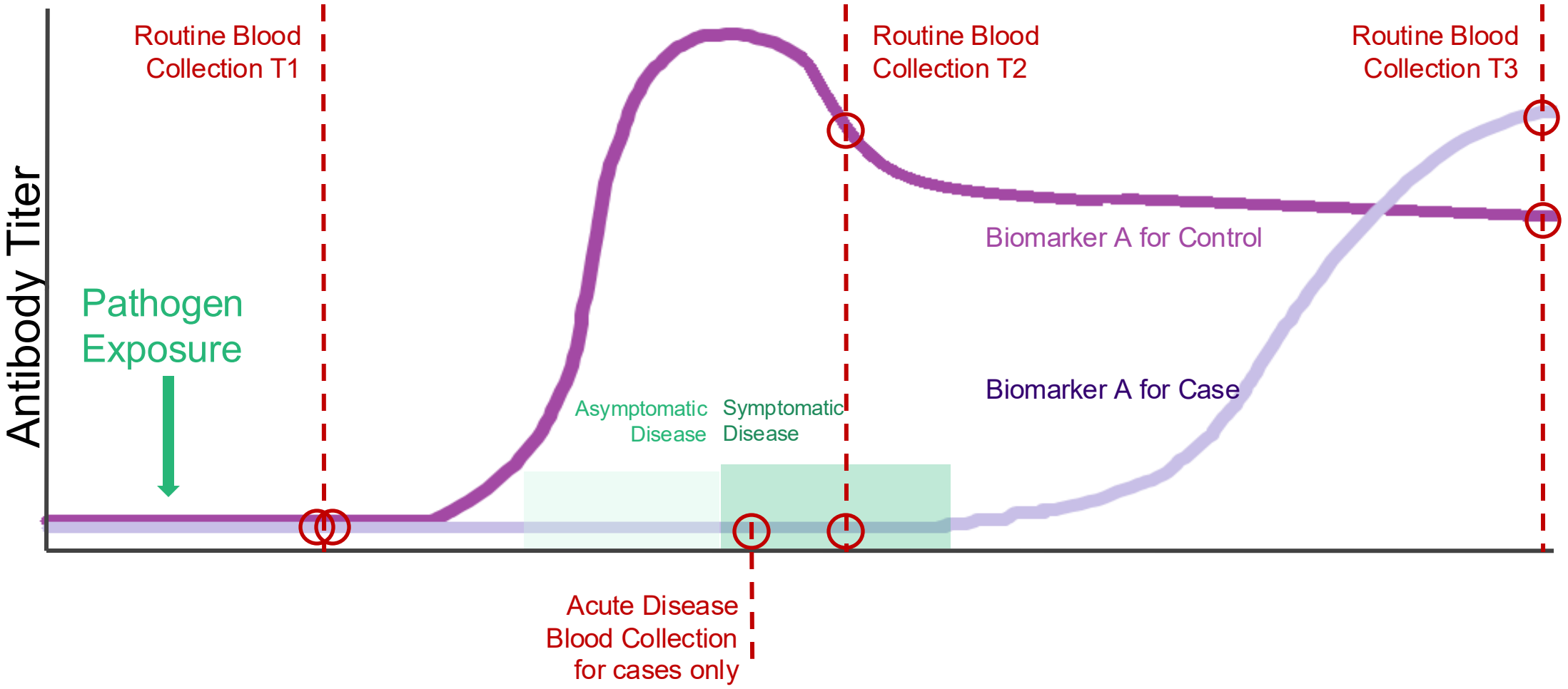
RECOMMENDED ANALYSIS

- Cases (events) for a given infectious disease outcome are matched to controls (non-events) from the risk set at the time of the event.
- Conditional matched logistic regression

ALTERATIVE CONSIDERED

- Case-cohort (shared control sub-cohort across all diseases)
- More efficient study design, but potentially ill-matched cases and controls

COP CASE-CONTROL ANALYSIS SCHEMATIC

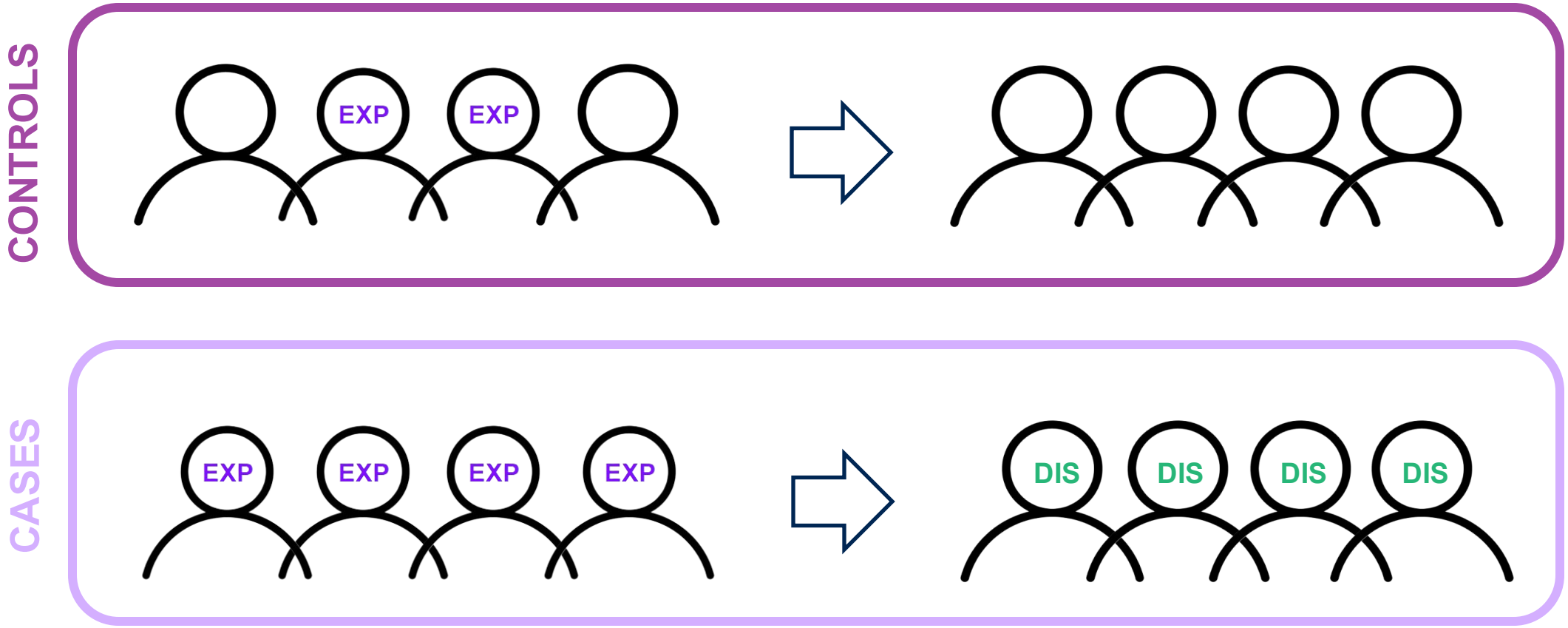


START CENTER

CONTROLS CONTAIN

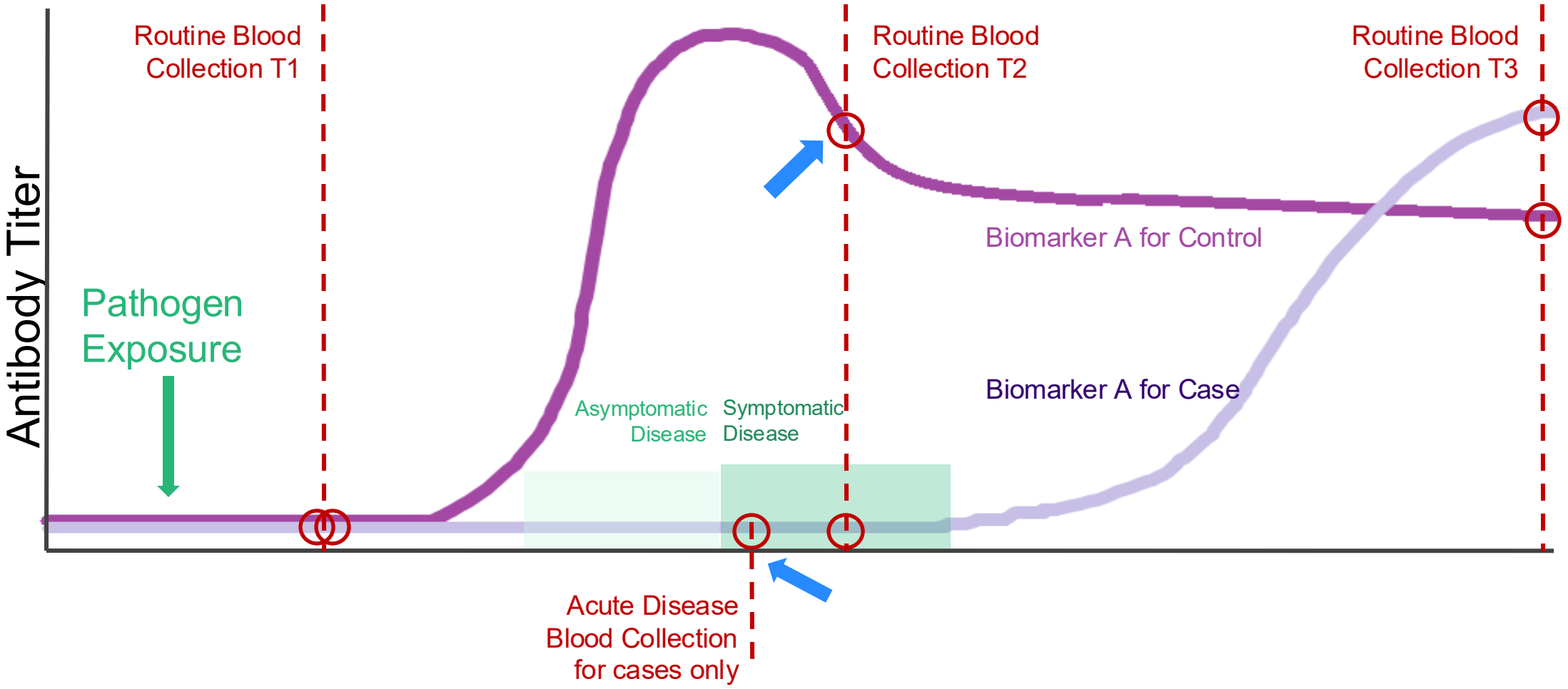
BOTH EXPOSED & NON-EXPOSED CHILDREN

ASSUMPTION



START CENTER

COP CASE-CONTROL ANALYSIS SCHEMATIC



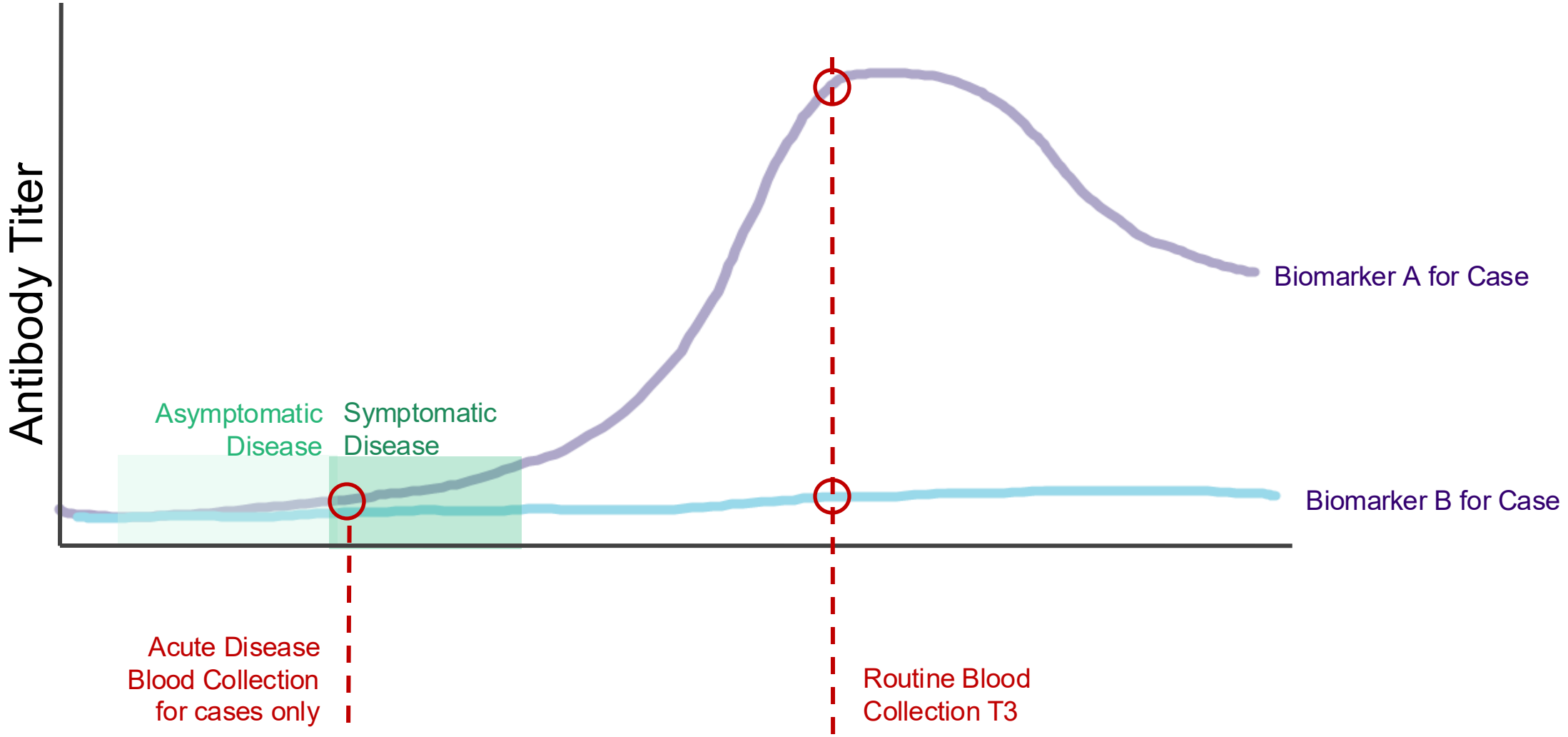
START CENTER

ALTERNATIVE (EFFICIENT) DESIGN PROPOSED DURING JUNE 5 MEETING

TEST-NEGATIVE or CASE-NEGATIVE

- Instead of doing a birth cohort with nested case control studies, you could do a disease surveillance cohort and test-negative design.
- Example of this (identifying COVID-19 COPs) here:
  - https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(23)00001-4/fulltext
  - https://www.medrxiv.org/content/10.1101/2022.10.24.22281399v2.full.pdf
- In this design, blood would be collected only from disease cases at the time of disease presentation (ideally early on in the disease).
  - Compare antibodies between cases (disease caused by pathogen of interest) and controls (disease caused by something other than pathogen of interest [if confirmed negative for disease of interest then this is a true "test negative")
- **Advantages:** This study design is much more efficient and practical as does not require long-term cohort and requires less blood sampling. Removes bias due to care-seeking (as cases and controls both sought care for the disease). Also lowers risk of differential exposure between cases and controls as both cases and controls have the disease (for example, diarrhea free controls may be inherently "less exposed" to the pathogens of interest). Can enroll far more cases since it's cross-sectional enabling additional analyses related to severity etc.
- **Disadvantages:** Not able to distinguish between first and subsequent infections which likely has implications for immune response; shared immune responses with other pathogens may be undetectable (although perhaps an advantage as you want COPs to be pathogen-specific); Only feasible for pathogens with short incubation period.
- Unrelated but potentially relevant pre-print: https://www.medrxiv.org/content/10.1101/2025.04.05.25325304v1

SECONDARY ANALYSES WITH EITHER DESIGN
(among cases only if blood sample available after disease)



START CENTER

POWER CALCULATIONS

What is the optimal number of cases and case-to-control ratio required to detect COP biomarkers with RR >=3 with at least 80% power?

POWER CALCULATIONS

What is the optimal number of cases and case-to-control ratio required to detect COP biomarkers with RR >=3 with at least 80% power?

Interpretation (Dichotomous Biomarker):

Samples with low protection levels of the biomarker are associated with >=3 times the risk of disease than samples with high protection levels of the biomarker

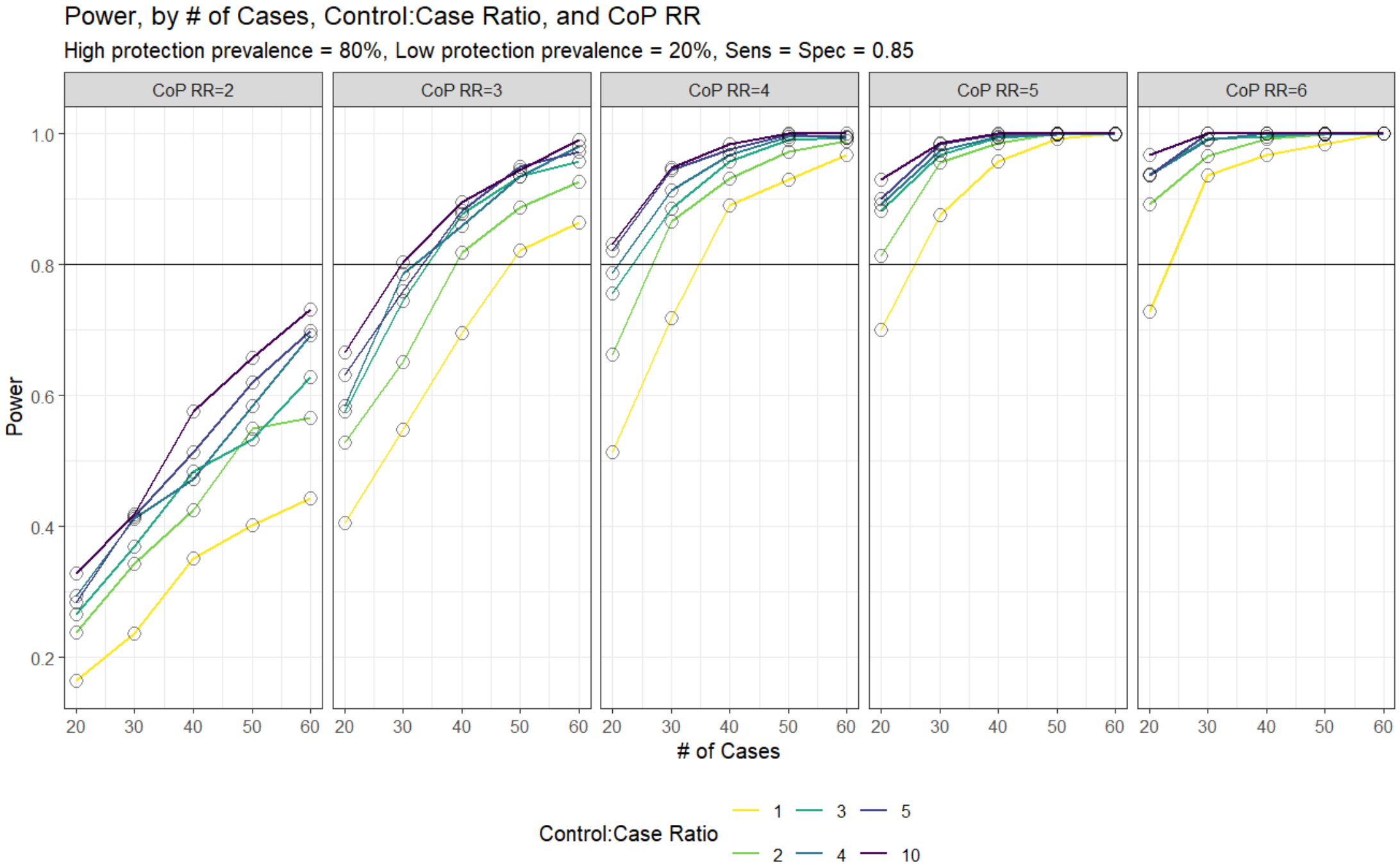
POWER CALCULATIONS

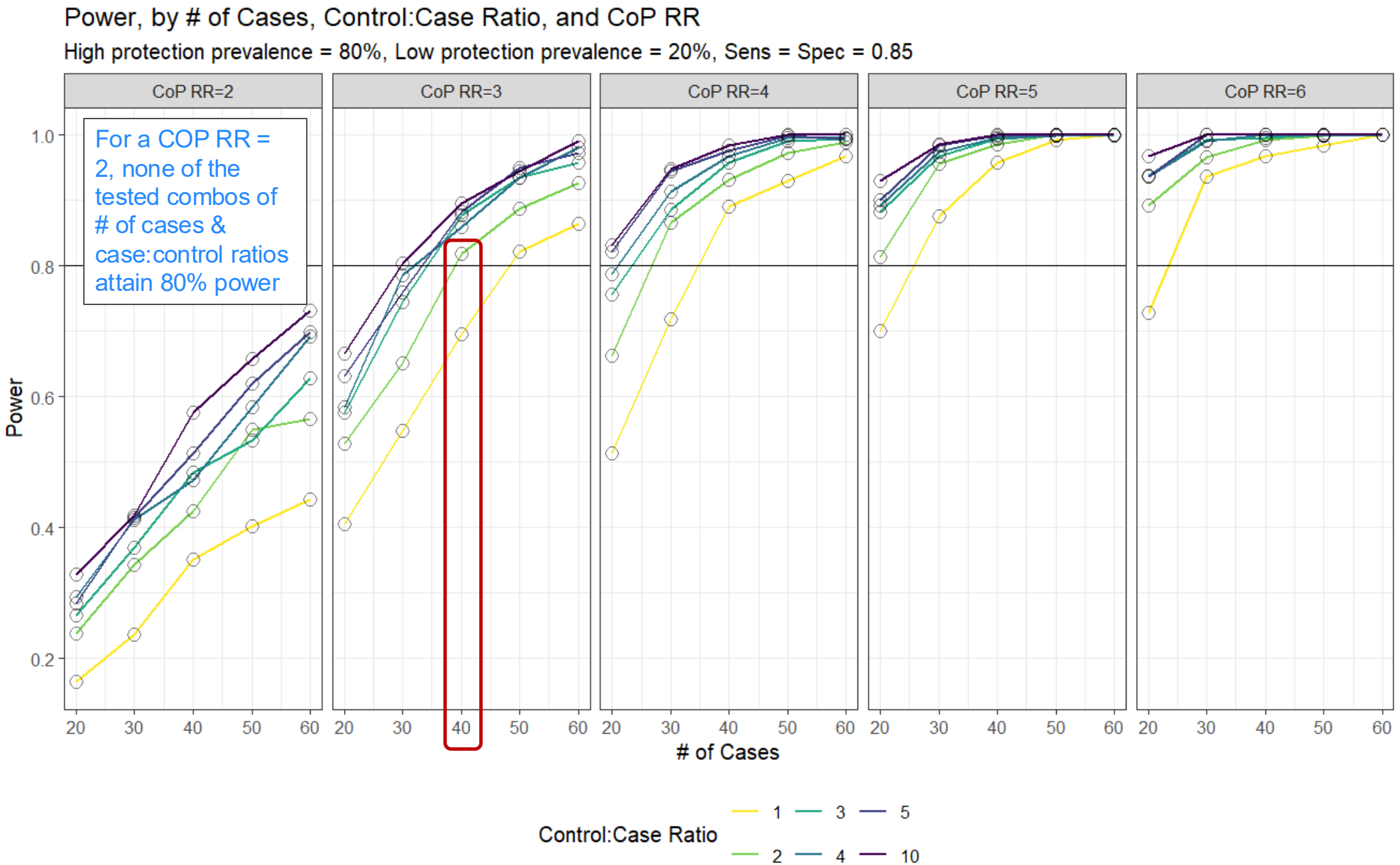
What is the optimal number of cases and case-to-control ratio required to detect COP biomarkers with RR >=3 with at least 80% power?

Method: Calculate power using methods devised for fixed-time correlates in clinical efficacy trials (CoRpower R package)

Requires data or assumptions on

- Prevalence of low protection and high protection groups (estimated at 20% low protection and 80% high protection)
- Measurement error of biomarker assay (estimated at 85% spec/sens)


Power, by # of Cases, Control:Case Ratio, and CoP RR
High protection prevalence = 80%, Low protection prevalence = 20%, Sens = Spec = 0.85
CoP RR=2
CoP RR=3
CoP RR=4
CoP RR=5
CoP RR=6
Power
1.0
0.8
0.6
0.4
0.2
20
30
40
50
60
# of Cases
Control:Case Ratio
1
2
3
4
5
10

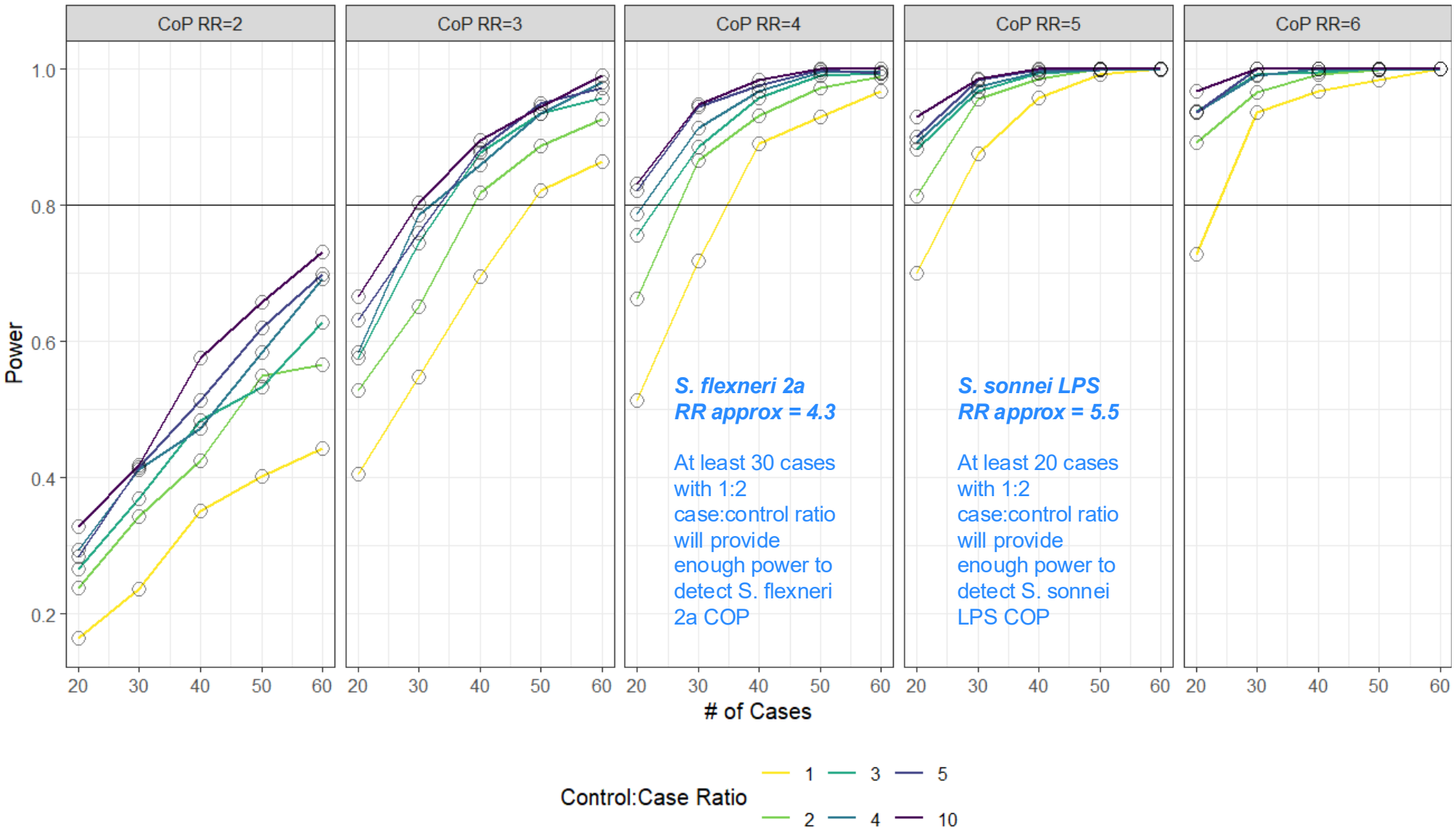

Power, by # of Cases, Control:Case Ratio, and CoP RR
High protection prevalence = 80%, Low protection prevalence = 20%, Sens = Spec = 0.85
CoP RR=2
CoP RR=3
CoP RR=4
CoP RR=5
CoP RR=6
For a COP RR = 2, none of the tested combos of # of cases & case:control ratios attain 80% power
Power
1.0
0.8
0.6
0.4
0.2
20
30
40
50
60
# of Cases
Control:Case Ratio
1
2
3
4
5
10

POWER CALCULATIONS

Recommendation: 40 cases with a minimum of 1:2 case:control ratio can identify potential COP biomarkers with RR >=3 with at least 80% power

Power calculations were estimated for dichotomous COPs; continuous COPs generally will require less power

CASE STUDY: SHIGELLA



Serum IgG antibodies to Shigella lipopolysaccharide antigens – a correlate of protection against shigellosis
Cohen 2019

*"Under these conditions of heavy natural exposure to Shigella, soldiers with "low" IgG titers to S. sonnei LPS at baseline were **5.5-fold** ($p = .0001$) more likely to develop S. sonnei shigellosis than soldiers with "high" titers. Similar analysis in S. flexneri 2a outbreaks showed odds ratios of 4.3 for ELISA IgG titers to S. flexneri 2a LPS." (Ref)*

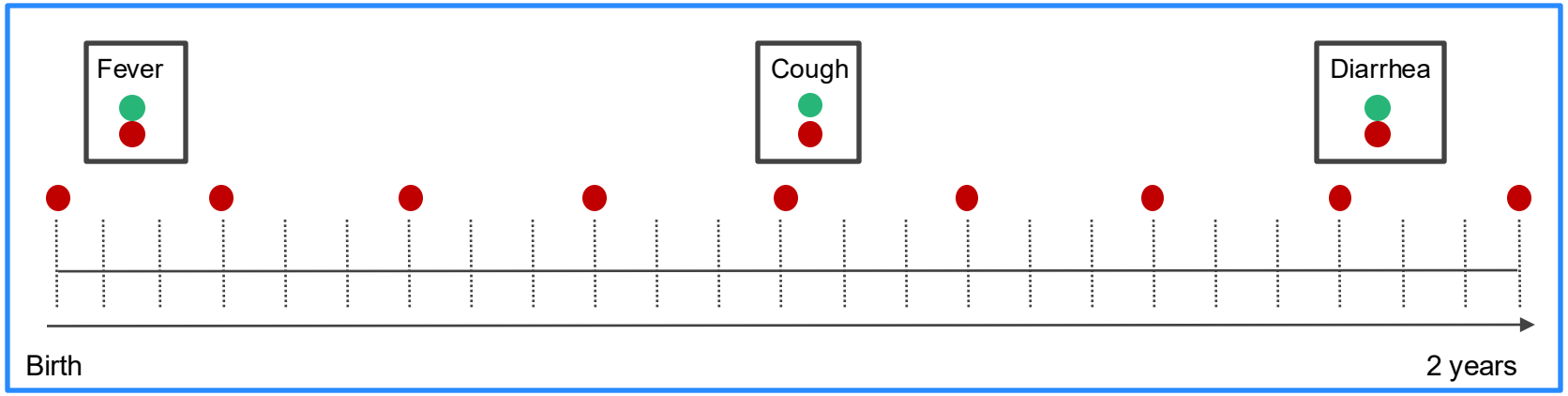
START CENTER

STUDY DESIGN: BIRTH COHORT

STUDY DESIGN RECOMMENDATIONS

Birth Cohort (Natural History Study)

2 years of follow-up and active disease surveillance and routine blood collection

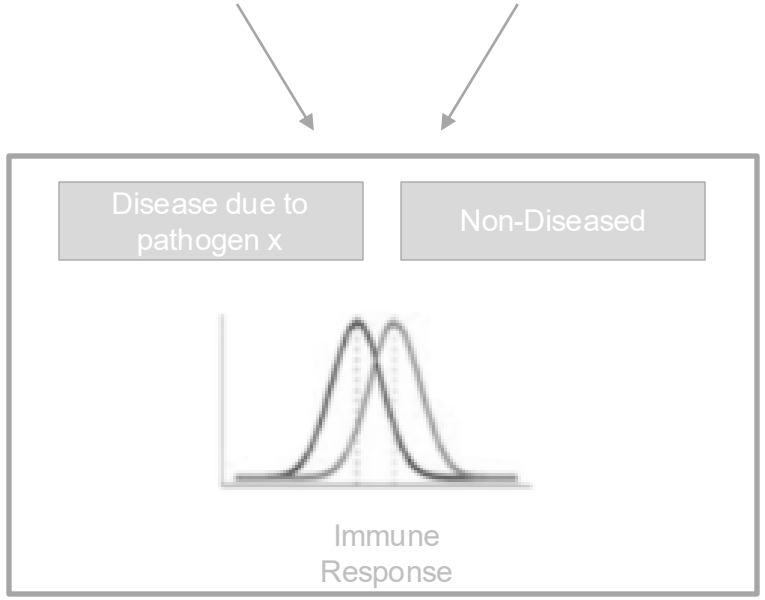


Routine study visit at study health facility

Blood sample for antibody testing

Relevant sample (stool, blood, NP swab) to diagnose disease etiology

Sick study visit



For each pathogen:

COP case-control analyses

From birth cohort, sample diseased (case) and non-diseased (control) children and compare blood samples to identify potential correlates of protection

START CENTER

BIRTH COHORT DESIGN

SIZE OF COHORT

GOAL

Identify correlates of protection for as many target pathogens as possible while keeping cohort size practical

CHALLENGE

Rare pathogens like cholera and iNTS require extremely large cohorts (up to 300,000+ in some geographies)

APPROACH

- Use incidence estimates from IHME and key studies to model cohort sizes for each pathogen and region
- Recommend a maximum feasible cohort size for inclusion

BIRTH COHORT DESIGN

SIZE OF COHORT

STUDY DESIGN RECOMMENDATIONS

2-Year birth cohort with routine blood sampling and disease surveillance with confirmatory diagnostic testing, including disease serogroup/subtype/sub-serotype

Stored blood samples (dried blood spots more feasible)

Nested case-control studies where cases are disease of interest and matched controls do not have the disease within +/- 2 months of the case occurrence

STUDY SIZING RECOMMENDATIONS

Sample size requirements for nested case control studies

40 cases of each disease
At least 2 controls per case
If require sufficient power for sub-serotype, then 20 cases of each sub-serotype of interest ae recommended

Sample size requirements for birth cohort

- Two separate birth cohorts each powered separately (to ensure one setting not driving the findings):
  - **3,800 children enrolled in South Asia**
  - **2,100 children in Sub-Saharan Africa**

If above is cost-prohibitive, we recommend a birth cohort with 1,900 children from SA and 1,100 from SSA

Likely to identify at least 40 disease cases of

Pertussis
Malaria (SSA only)
Cryptosporidium
ETEC
Shigella
Norovirus
Rotavirus
Campylobacter
Adenovirus
RSV

Unlikely to identify at least 40 disease cases of

iNTS
Cholera

BIRTH COHORT SIZING DATA & METHODS

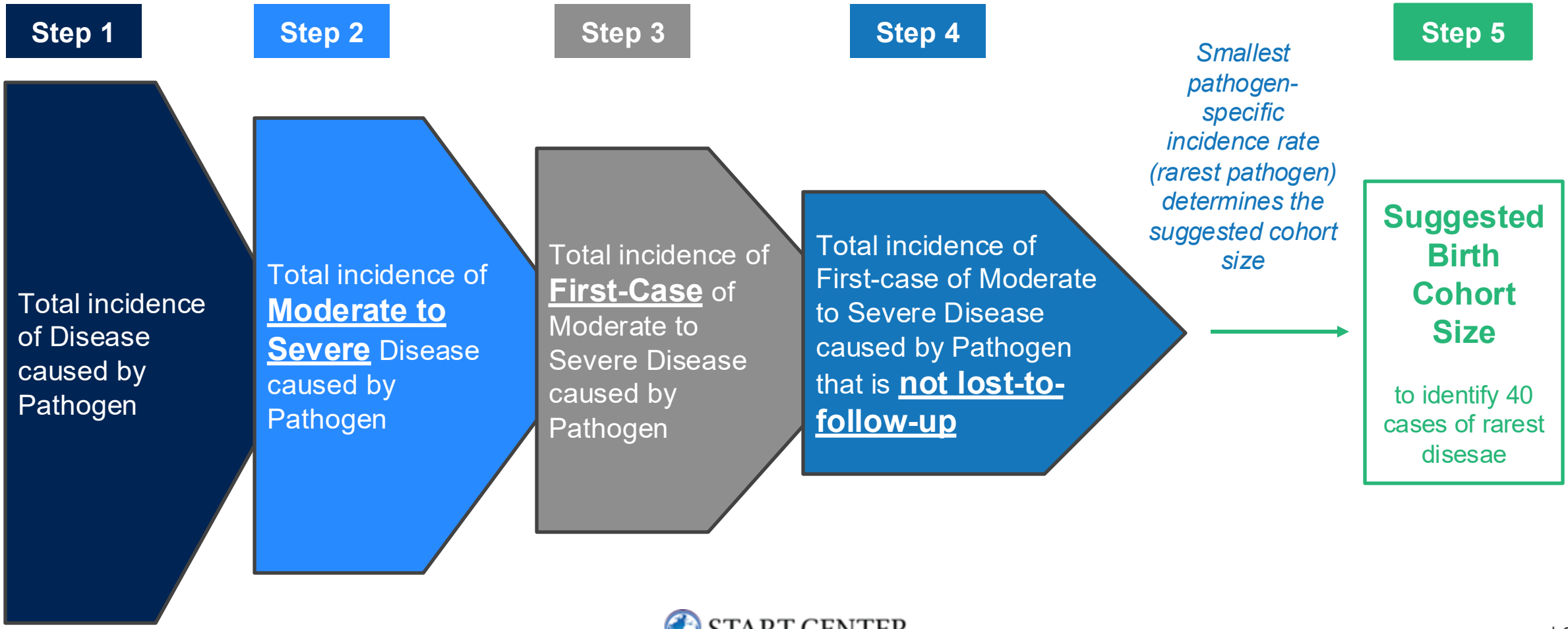
Study Parameters

- Birth cohort study with 2-year follow up
- Regions: South Asia and sub-Saharan Africa
- Target # of cases per pathogen: 40

Incidence Rates	
Total disease incidence rates based on IHME GBD 2021 with literature comparisons for RSV, pertussis, and diarrheal pathogens	IHME incidence estimates include age-specific (0-28 days, 1-5 mo, 6-11 mo, 12 -24mo) and geography-specific (SSA & SA) incidence rates. For some pathogens, IHME incidence had to be calculated by multiplying the YLD PAF by the overarching cause of disease (e.g. YLD% for Shigella x total diarrheal incidence. Literature comparisons came from MAL-ED and other key studies
Adjusted total incidence of disease to incidence of moderate-severe disease	Moderate to severe disease is more likely than mild disease to be averted by a vaccine therefore a COP is more likely to be identified in this subset of cases.
Adjusted to estimate first disease only	For common infections, disease may occur more than once and second infections may solicit different immune response. Powering for first disease provides cleanest analysis but enables secondary analyses the include secondary and third instances of disease

BIRTH COHORT DESIGN

SIZE OF COHORT



BIRTH COHORT SIZE

ADJUSTMENTS

Pathogen	Moderate-to-Severe adjustment %	First disease adjustment	Loss-to-follow-up
Cholera	20% (18% - 25%)	100%	10%
iNTS	0%	100% (assumed rare enough)	
Malaria	30% (20% - 40%)	13%	
RSV	20% (15% - 25%)	67%	
Pertussis	30% (25% - 35%)	100% (assumed rare enough)	
Cryptosporidium	20%(18% - 25%)	86%	
ETEC	20%(18% - 25%)	73%	
Shigella	20% (18% - 25%)	67%	
Norovirus	20% (18% - 25%)	84%	
Rotavirus	20% (18% - 25%)	79%	
Campylobacter	20% (18% - 25%)	77%	
Adenovirus	20% (18% - 25%)	60%	

BIRTH COHORT SIZE

INCIDENCE SOURCES

	Source of incidence rate (% etiology of disease or directly estimated)	Geographies used	Age-specific incidence
Pertussis	IHME: Directly estimated by IHME	SSA & SA	0-2y incidence (aggregated from 0-28d, 28d-6m, 6m-1y, 1y-2y incidence from IHME)
iNTS	IHME: Directly estimated by IHME	SSA & SA	
Malaria	IHME: Directly estimated by IHME	SSA & SA	
Cholera	IHME: Inc = % YLD x diarrhea inc	SSA & SA	
Cryptosporidium	IHME: Inc = % YLD x diarrhea inc	SSA & SA	
ETEC	IHME: Inc = % YLD x diarrhea inc	SSA & SA	
Shigella	IHME: Inc = % YLD x diarrhea inc	SSA & SA	
Norovirus	IHME: Inc = % YLD x diarrhea inc	SSA & SA	
Rotavirus	IHME: Inc = % YLD x diarrhea inc	SSA & SA	
Campylobacter	IHME: Inc = % YLD x diarrhea inc	SSA & SA	
Adenovirus	IHME: Inc = % YLD x diarrhea inc	SSA & SA	
RSV	Non-IHME: 2019 Systematic analysis	137 LMICs	0-12 m (assumed equal to 0-2y incidence)

BIRTH COHORT SIZE

SUB-SAHARAN AFRICA

Pathogen	Cohort Size
Cholera	83,061
iNTS	13,350
Pertussis	2,042
Malaria	1,858
RSV	1,491
Campylobacter	1,430
ETEC	1,129
Norovirus	1,109
Cryptosporidium	955
Adenovirus	830
Shigella	822
Rotavirus	598

Recommended Max Cohort Size = 2,100

Pathogens included

START CENTER

BIRTH COHORT SIZE

SOUTH ASIA

Pathogen	Cohort Size
Cholera	347,878
iNTS	347,041
Malaria	135,747
Pertussis	3,714
Cryptosporidium	2,609
RSV	1,491
ETEC	1,704
Shigella	1,462
Norovirus	1,269
Rotavirus	1,195
Campylobacter	848
Adenovirus	590

Recommended Max Cohort Size = 3,800

Pathogens included

DETECTABLE RRs WITH RECOMMENDED COHORT SIZES

SUB-SAHARAN AFRICA = 2,100 children

Pathogen	Total Cases	RR Detectable
Pertussis	41	3.0
Malaria	45	2.9
Campylobacter	59	2.4
ETEC	74	2.2
Norovirus	76	2.2
Cryptosporidium	88	2.1
Adenovirus	101	1.9
Shigella	102	1.9
RSV	107	1.9
Rotavirus	140	1.7

SOUTH ASIA = 3,800 children

Pathogen	Total Cases	RR Detectable
Pertussis	41	2.9
Cryptosporidium	58	2.4
ETEC	89	2.1
Shigella	103	1.9
RSV	107	1.9
Norovirus	119	1.8
Rotavirus	127	1.8
Campylobacter	179	1.6
Adenovirus	258	1.5

BIRTH COHORT SIZE

SHIGELLA BY SEROGROUP/S. FLEXNERI SUBSEROTYPES

Cohort Size = 3,800 in South Asia

Shigella Serogroup	% of Total Shigella	Expected Cases
Shigella overall	100%	103
S. flexneri	66%	67
S. sonnei	24%	24
S. boydii	5.4%	5
S. dysenteriae	5.0%	5

Shigella Subserotype	% of Total Shigella	Expected Cases
Shigella flexneri overall	66%	67
S. flexneri 2a	20%	21
S. flexneri 6	11%	11
S. flexneri 3a	9.4%	9
S. flexneri 1b	7.5%	7
S. flexneri 4a	2.9%	2

**Proportions of Shigella serogroups and serotypes/subserotypes based on Livio 2014*

BIRTH COHORT SIZE

SHIGELLA BY SEROGROUP/S. FLEXNERI SUBSEROTYPES

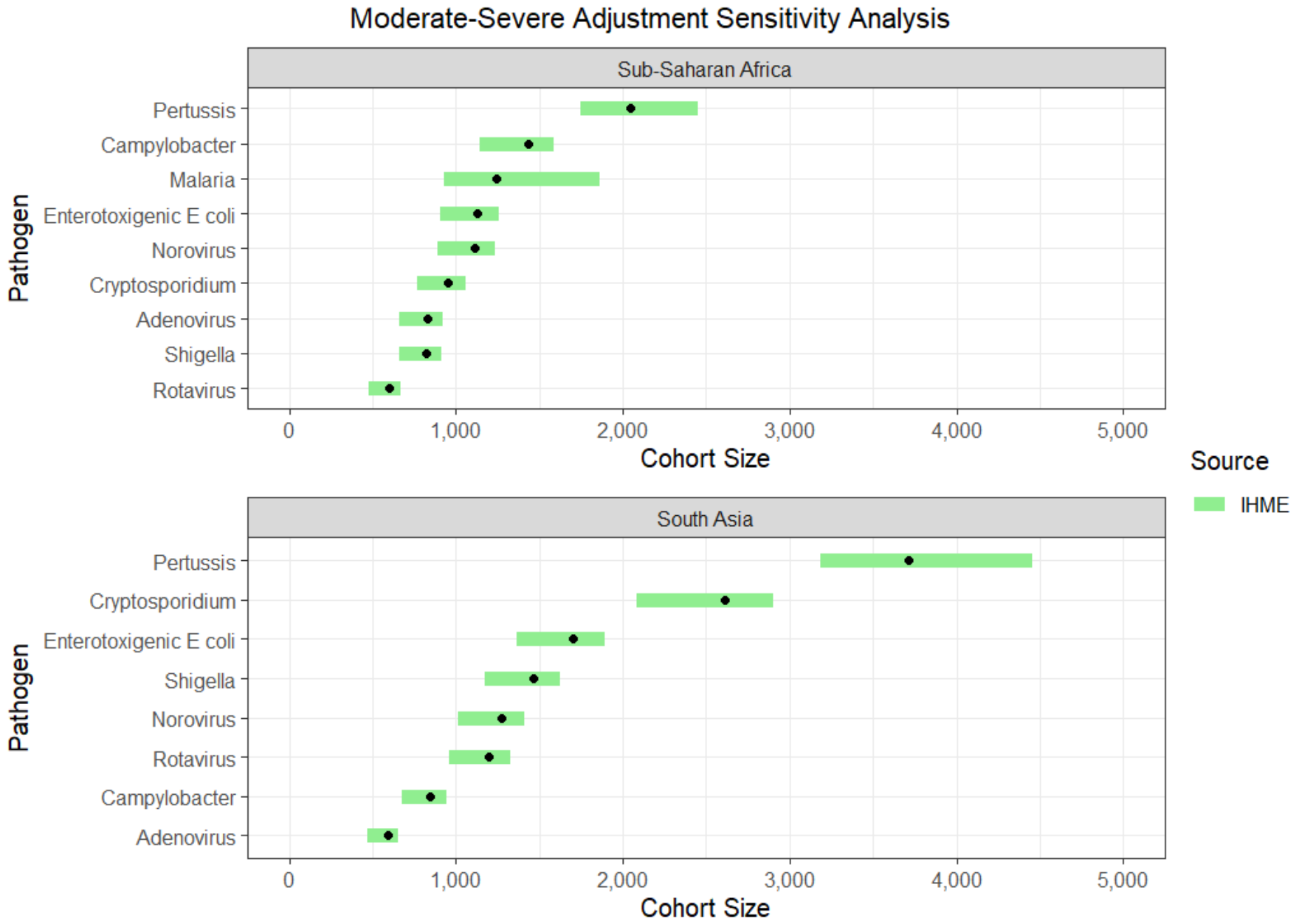
Cohort Size = 2,100 in SS Africa

Shigella Serogroup	% of Total Shigella	Expected Cases
Shigella overall	100%	102
S. flexneri	66%	67
S. sonnei	24%	24
S. boydii	5.4%	6
S. dysenteriae	5.0%	5

Shigella Subserotype	% of Total Shigella	Expected Cases
Shigella flexneri overall	66%	67
S. flexneri 2a	20%	21
S. flexneri 6	11%	11
S. flexneri 3a	9.4%	10
S. flexneri 1b	7.5%	8
S. flexneri 4a	2.9%	3

**Proportions of Shigella serogroups and serotypes/subserotypes based on Livio 2014*

MODERATE-SEVERE ADJUSTMENTS: SENSITIVITY ANALYSIS



Source: IHME Estimates

Moderate-Severe adjustments:

- Diarrheal pathogens: 18%, 20%, and 25%
- Pertussis: 25%, 30%, 35%
- Malaria: 20%, 30%, 40%

KEY STUDY DESIGN CONSIDERATIONS

KEY STUDY DESIGN CONSIDERATIONS

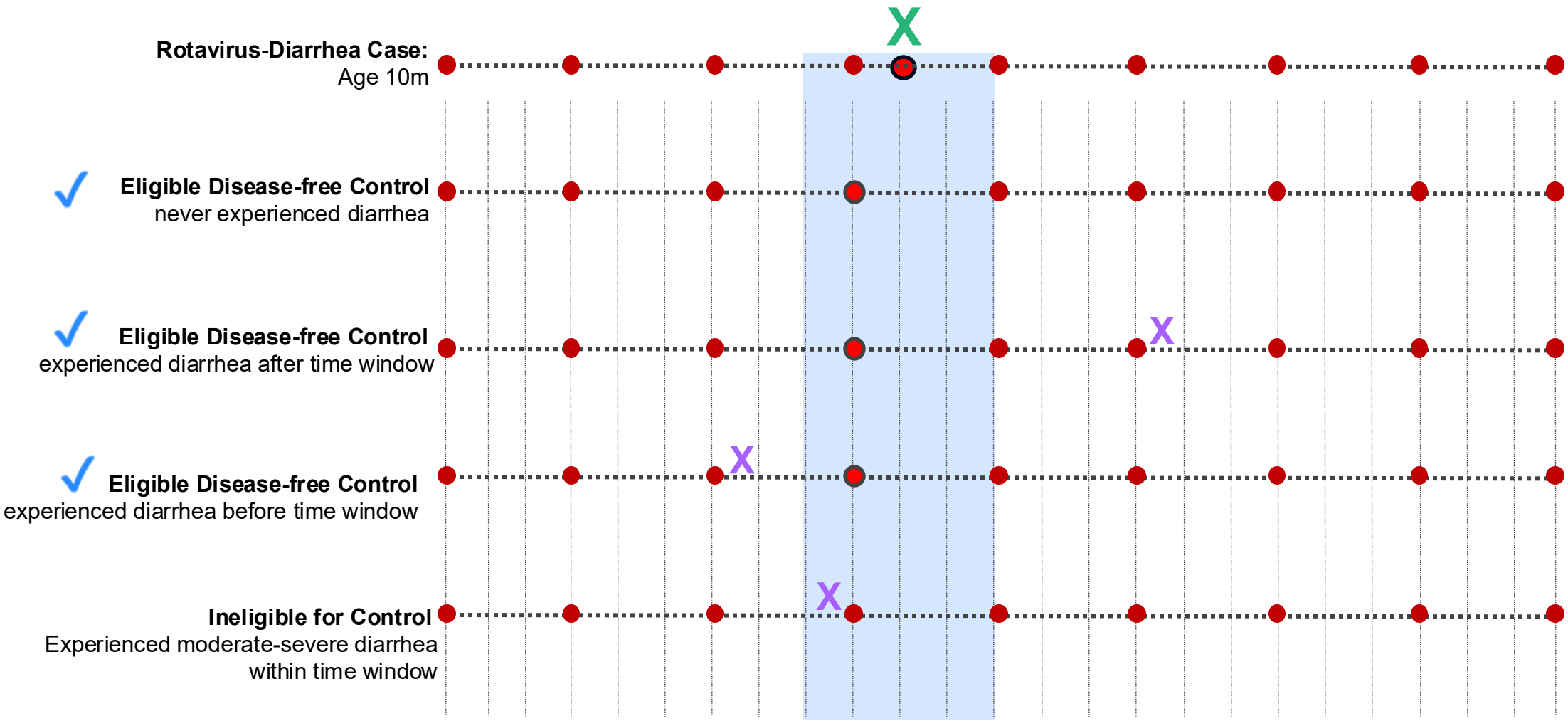
1. Selection of appropriate controls
2. Addressing confounding
3. IHME vs Non-IHME incidence
4. Blood sampling
5. Moderate-to-severe disease definitions

SELECTION OF CONTROLS

RECOMMENDATION: Exclude controls experiencing **moderate-severe** symptoms two months before or after case

RATIONALE

Controls with moderate-severe symptoms likely have systemically elevated immune responses compared to controls without moderate-severe symptoms



SHOULD CHILDREN WITH MILD SYMPTOMS 2-MONTHS BEFORE/AFTER THE CASE BE ELIGIBLE AS MATCHED CONTROLS?

Advantages of including controls with mild symptoms	Advantages of excluding controls with any symptoms
Minimizes selection bias and ensures representativeness	Clear case-control distinction, maximizes phenotypic separation
Avoids biased estimates of etiology, through standardized training and clinical assessment	Lower risk of including pre-cases with early disease
Accurate pathogen prevalence as including symptomatic controls helps reflect true prevalence	Stronger effect sizes, may enhance detection of pathogen associations
Mimics cohort studies, aligns with gold standard epidemiological principles	Simplified analysis, less concern about symptom gradients
Avoids intermediate phenotype bias, prevents overestimation of pathogen associations	Clearer distinction between healthy and disease state, especially where strict definition is desired

Should Controls With Respiratory Symptoms Be Excluded From Case-Control Studies of Pneumonia Etiology? Reflections From the PERCH Study

Higdon 2017

"The PERCH study demonstrates that including controls with mild symptoms, while requiring careful implementation, provides less biased estimates of pneumonia etiology by maintaining population representativeness. The trade-off favors inclusivity when robust case definitions and follow-up procedures are in place"

ADDRESSING CONFOUNDING

Proposed approaches to addressing confounding

01 Study Design: Matching

02 Analysis: Collect data on confounders and adjust in regression

POTENTIAL CONFOUNDERS

- Age
- Study site/region
- Vaccination status
- Sociodemographic status
- Malnutrition

BIRTH COHORT SIZE

NON-IHME INCIDENCE SOURCES

	Source	Geographies used	Age range of incidence
Cholera			
iNTS			
Malaria			
RSV	2019 Systematic analysis	137 LMICs	0-12 mo
Pertussis	SAMIPS study	Zambia	0-14 weeks
Cryptosporidium	MAL-ED	6 LMICS*	0-24 mo
ETEC	MAL-ED	6 LMICS*	0-24 mo
Shigella	MAL-ED	6 LMICS*	0-24 mo
Norovirus	MAL-ED	6 LMICS*	0-24 mo
Rotavirus	MAL-ED	6 LMICS*	0-24 mo
Campylobacter	MAL-ED	6 LMICS*	0-24 mo
Adenovirus	MAL-ED	6 LMICS*	0-24 mo

**For the MAL-ED study, we averaged incidence rates from Bangladesh, India, Nepal, and Pakistan*

BIRTH COHORT SIZE

MAL-ED INCIDENCE ESTIMATES

Table S2. Attributable incidence of diarrhoea by quantitative PCR in MAL-ED by site and overall.

	Dhaka, Bangladesh	Vellore, India	Bhaktapur, Nepal	Naushero Feroze, Pakistan	Venda, South Africa*	Haydom, Tanzania	Fortaleza, Brazil*	Loreto, Peru*	Overall
Overall incidence	361·9	211·5	233·5	632·1	62·2	128·5	50·9	449	273·8
Shigella	65·2 (57·4–76·9)	27·9 (21·0–35·8)	18·6 (14·7–24·1)	33·6 (26·3–44·4)	4·2 (1·5–8·1)	9·3 (1·3–17·8)	6·8 (3·4–11·5)	42·3 (32·5–53·0)	26·1 (23·8–29·9)
Sapovirus	28·6 (17·9–39·2)	20·3 (13·5–27·0)	23·2 (16·8–29·1)	36·6 (20·4–54·2)	3·5 (1·0–6·0)	11·2 (4·9–18·5)	3·7 (1·4–6·6)	54·0 (43·3–67·0)	22·8 (18·9–27·5)
Rotavirus	57·6 (50·4–66·3)	19·5 (14·8–26·1)	23·3 (19·3–27·8)	25·9 (19·7–32·4)	2·5 (0·5–4·9)	13·4 (7·6–21·1)	1·5 (0·0–3·3)	19·5 (13·9–25·3)	20·7 (18·8–23·0)
Adenovirus 40/41	86·5 (73·8–105·6)	12·8 (6·0–19·9)	4·3 (0·7–8·6)	7·3 (0·0–18·8)	2·3 (0·1–4·8)	7·1 (2·0–13·5)	3·0 (0·6–6·0)	32·4 (23·1–44·4)	19·0 (16·8–23·0)
ETEC	55·5 (44·2–68·7)	17·3 (12·7–25·4)	15·2 (10·4–20·7)	17·6 (10·7–31·0)	1·9 (0·4–4·3)	22·5 (13·1–36·5)	2·3 (0·7–5·2)	16·8 (9·7–30·3)	18·8 (16·5–23·8)
Norovirus	16·4 (9·7–26·3)	7·3 (2·4–12·3)	16·5 (12·3–22·7)	16·1 (7·4–32·9)	3·7 (1·5–8·5)	14·2 (7·3–23·8)	6·3 (3·0–10·4)	44·8 (35·2–61·2)	15·4 (13·5–20·1)
Astrovirus	18·8 (8·3–28·9)	13·7 (8·6–21·2)	9·2 (5·8–13·1)	28·7 (18·2–43·9)	2·8 (0·8–6·0)	4·5 (0·8–9·1)	1·6 (0·4–3·7)	39·5 (31·3–53·6)	15·0 (12·0–19·5)
C. jejuni/coli	14·2 (3·8–29·8)	11·1 (5·1–17·2)	8·2 (2·1–15·7)	14·1 (1·3–31·1)	3·2 (0·5–6·6)	6·9 (0·0–17·9)	2·4 (0·5–6·6)	37·6 (24·5–51·9)	12·1 (8·5–17·2)
Cryptosporidium	6·9 (2·3–11·8)	5·6 (2·7–12·4)	4·1 (1·5–6·5)	8·8 (4·0–18·6)	0·2 (0·0–1·2)	2·2 (0·0–7·1)	0·3 (0·0–1·1)	18·2 (10·8–26·7)	5·8 (4·3–8·3)
tEPEC	4·6 (0·2–11·4)	5·9 (1·0–12·7)	3·7 (1·0–7·5)	15·6 (2·0–34·8)	0·6 (00–3·2)	4·2 (0·1–11·4)	0·1 (0·0–0·9)	6·5 (0·8–13·3)	5·4 (2·8–9·3)
EAEC		0·2 (0·0–2·8)	0·4 (0·0–3·4)	2·7 (0·3–15·1)	2·2 (0·0–8·6)	9·7 (0·7–28·9)	4·0 (0·2–9·8)	1·4 (0·1–8·2)	2·5 (1·1–6·0)
Giardia	0·9 (0·0–5·5)	0·7 (0·0–4·2)	0·2 (0·0–1·5)					8·5 (2·1–18·8)	1·2 (0·4–2·8)
Plesiomonas	0·8 (0·0–6·5)	0·7 (0·0–4·6)	0·2 (0·0–0·8)			1·9 (0·0–7·3)		2·6 (0·0–9·2)	0·7 (0·1–2·2)
aEPEC	1·0 (0·1–2·4)	0·1 (0·0–0·5)	0·9 (0·1–3·5)					2·7 (0·5–7·6)	0·6 (0·1–1·5)
Isospora	0·3 (0·0–1·9)		0·2 (0·0–1·2)				0·9 (0·0–2·8)	2·9 (0·4–7·7)	0·5 (0·1–1·1)
V. cholerea	1·1 (0·2–2·6)	1·3 (0·2–2·8)							0·3 (0·1–0·6)
Cyclospora	0·9 (0·0–3·0)		0·6 (0·0–2·5)					0·9 (0·0–2·6)	0·3 (0·1–0·7)
E. bieneusi	0·3 (0·0–2·7)		0·1 (0·0–1·7)	0·3 (0·0–2·1)				1·7 (0·1–10·8)	0·3 (0·0–1·8)
E. histolytica	1·6 (0·3–3·5)	0·6 (0·0–1·7)							0·3 (0·1–0·5)
Salmonella		0·2 (0·0–1·5)			0·3 (0·0–1·7)	0·4 (0·0–1·7)	0·5 (0·0–1·5)		0·2 (0·0–0·7)
Strongyloides								1·4 (0·0–4·5)	0·2 (0·0–0·6)
Ancylostoma								1·4 (0·3–2·9)	0·2 (0·1–0·4)
H. pylori					0·1 (0·0–0·9)	0·4 (0·0–3·5)			0·1 (0·0–1·1)
Trichuris	0·2 (0·0–2·5)						0·1 (0·0–0·8)		0·0 (0·0–0·4)
E. intestinalis	0·1 (0·0–0·5)								0·0 (0·0–0·3)

Estimates are per 100 child-years (95% CIs) and are ordered by overall attributable incidence. All pathogens with a point estimate greater than 0 for at least one site or overall are shown. EAEC=enteroaggregative E
aEPEC=atypical enteropathogenic E. coli. tEPEC=typical enteropathogenic E. coli. ETEC=enterotoxigenic E. coli

BIRTH COHORT SIZE

NON-IHME INCIDENCE SOURCES

PERTUSSIS

3). The overall pertussis incidence was 2.4 cases per 1000 infant-months (95% confidence interval [CI], 1.2–4.2). Using the maximum severity score for each infected infant, only 1 infant qualified as "severe pertussis" via MPS criteria (incidence rate, 0.2 cases/1000 infant-months), meaning that 90% of infected infants had nonsevere pertussis (incidence rate, 2.1 cases per 1000 infant-months). Parenthetically, we note that the single case of severe pertussis also met the CDC screening case definition for pertussis.

Table 3. Incidence of Severe and Nonsevere Infant Pertussis

Pertussis	No. of Infants	Person-time, months	Incidence Rate per 1000 Person-months (95% CI)	Cumulative Incidence per 1000 Infants (95% CI)
All pertussis	10	4254	2.4 (1.2–4.2)	5.1 (2.6–9.0)
Nonsevere pertussis	9	4254	2.1 (1.0–3.9)	4.5 (2.2–8.3)
Severe pertussis	1	4254	0.2 (.1–1.6)	0.5 (.3–2.5)

Abbreviation: CI, confidence interval.

Gill CJ, Mwananyanda L, MacLeod W, et al. Incidence of Severe and Nonsevere Pertussis Among HIV-Exposed and -Unexposed Zambian Infants Through 14 Weeks of Age: Results From the Southern Africa Mother Infant Pertussis Study (SAMIPS), a Longitudinal Birth Cohort Study.

RSV

Table 1. Incidence and number of episodes of RSV-associated acute lower respiratory infection in children younger than 5 years in 2019, by World Bank income regions and development status

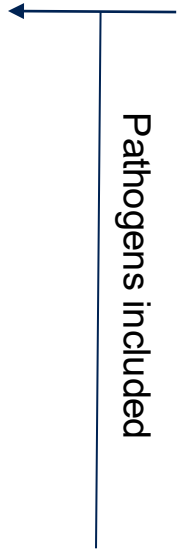
	Low income	Lower-middle income	Upper-middle income	High income	Developing countries	Industrialised countries	Global*
RSV-associated acute lower respiratory infection							
0–12 months†							
Studies	5	8	6	5	19	5	24
Incidence rate	78·3 (43·2–142·2)	111·2 (81·7–151·2)	108·8 (48·6–243·7)	38·5 (21·6–68·8)	101·0 (72·5–140·6)	38·5 (21·6–68·8)	94·6 (70·8–131·6)
Number of episodes	1 902 000 (1 048 000–3 453 000)	6 969 000 (5 123 000–9 480 000)	3 885 000 (1 735 000–8 698 000)	515 000 (288 000–920 000)	12 401 000 (8 907 000–17 267 000)	510 000 (286 000–911 000)	12 875 000 (9 635 000–17 909 000)

Per 1,000 child-years
Li, You, et al. "Global, Regional, and National Disease Burden Estimates of Acute Lower Respiratory Infections Due to Respiratory Syncytial Virus in Children Younger than 5 Years in 2019: A Systematic Analysis."

BIRTH COHORT SIZE

SUB-SAHARAN AFRICA - COMPARISON

Pathogen	Non-IHME	IHME
Cholera		83,061
RSV	1,491	20,196
iNTS		13,350
Pertussis	2,646	2,042
Malaria		1,858
Campylobacter		1,430
ETEC	1,248	1,129
Norovirus	1,478	1,109
Cryptosporidium	10,767	955
Adenovirus	3,940	830
Shigella	1,228	822
Rotavirus	1,769	598

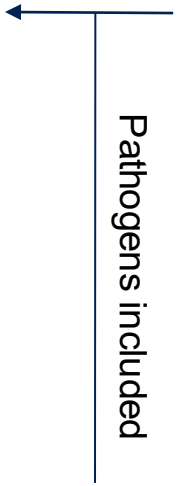


START CENTER

BIRTH COHORT SIZE

SOUTH ASIA - COMPARISON

Pathogen	Non-IHME	IHME
Cholera	9,259	347,878
iNTS		347,041
Malaria		135,747
RSV	1,491	14,110
Pertussis	2,646	3,714
Cryptosporidium	2,035	2,609
ETEC	577	1,704
Shigella	457	1,462
Norovirus	940	1,269
Rotavirus	445	1,195
Campylobacter		848
Adenovirus	668	590



BLOOD SAMPLING TECHNIQUES

KEY FEATURES OF DRIED BLOOD SPOT VS VENIPUNCTURE

Features	Dried Blood Spot	Serum Blood
Blood volume	Small volume (a few drops from a finger prick)	Larger volume (e.g., 1–10 mL of venous blood
Sampling stability	Stable at room temperature	Requires cold storage
Collection	Minimally invasive	Invasive (venipuncture)
Quantification	Less precise	Precise
Cost	Lower	Higher

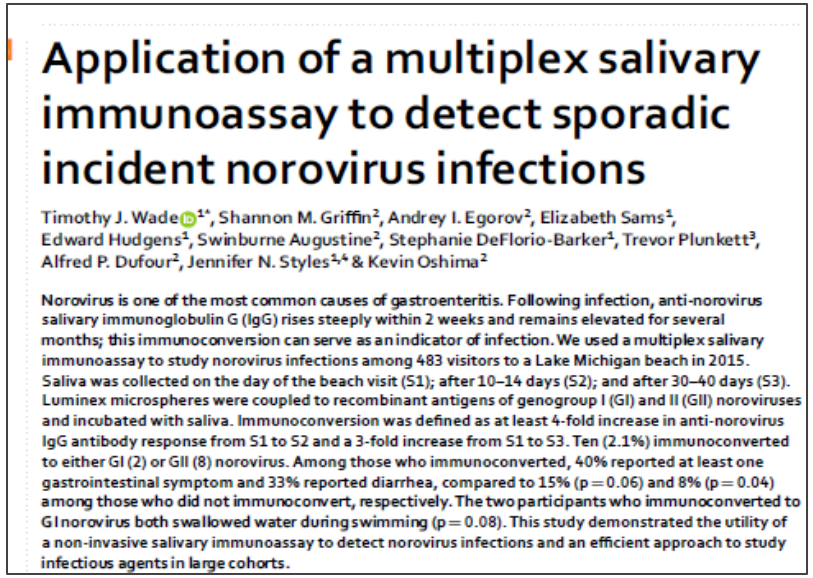
What a drop can do: Dried blood spots as a minimally invasive method for integrating biomarkers into population-based research

McDade 2007

"We conclude that for many biomarkers, DBS sampling provides a viable alternative to using venipuncture, particularly as the long list of analytes that can be quantified in blood spot samples grows."

SYSTEMIC RESPONSES FROM PATHOGENS

Pathogen	Systemic response
Cholera	✓✓
iNTS	
Malaria	
Pertussis	
RSV	✓ *Primary response in local respiratory tract*
Adenovirus	✓ *Primary response is mucosal*
Shigella	
Campylobacter	
Rotavirus	
Cryptosporidium	? *Primary response is mucosal*
ETEC	
Norovirus	



Application of a multiplex salivary immunoassay to detect sporadic incident norovirus infections

Timothy J. Wade[1*], Shannon M. Griffin[2], Andrey I. Egorov[2], Elizabeth Sams[1], Edward Hudgens[1], Swinburne Augustine[2], Stephanie DeFlorio-Barker[1], Trevor Plunkett[3], Alfred P. Dufour[2], Jennifer N. Styles[1,4] & Kevin Oshima[2]

Norovirus is one of the most common causes of gastroenteritis. Following infection, anti-norovirus salivary immunoglobulin G (IgG) rises steeply within 2 weeks and remains elevated for several months; this immunoconversion can serve as an indicator of infection. We used a multiplex salivary immunoassay to study norovirus infections among 483 visitors to a Lake Michigan beach in 2015. Saliva was collected on the day of the beach visit (S1); after 10–14 days (S2); and after 30–40 days (S3). Luminex microspheres were coupled to recombinant antigens of genogroup I (GI) and II (GII) noroviruses and incubated with saliva. Immunoconversion was defined as at least 4-fold increase in anti-norovirus IgG antibody response from S1 to S2 and a 3-fold increase from S1 to S3. Ten (2.1%) immunoconverted to either GI (2) or GII (8) norovirus. Among those who immunoconverted, 40% reported at least one gastrointestinal symptom and 33% reported diarrhea, compared to 15% ($p = 0.06$) and 8% ($p = 0.04$) among those who did not immunoconvert, respectively. The two participants who immunoconverted to GI norovirus both swallowed water during swimming ($p = 0.08$). This study demonstrated the utility of a non-invasive salivary immunoassay to detect norovirus infections and an efficient approach to study infectious agents in large cohorts.

✓✓	Strong Systemic Response
✓	Systemic response can occur
?	Limited systemic response

CASE DEFINITIONS

Pathogen	Clinical Case Definition	Etiologic confirmation
Cholera	**Diarrhea**: 3 or more abnormally loose or watery stool with or without visible blood **Moderate or severe**: Diarrhea as defined above plus one or more of: dehydration, visible blood in stool, and/or recommended to be hospitalized	Microbiologic culture for bacteria, ELISAs for the viruses and parasites, quantitative PCR using "attributable" thresholds
Cryptosporidium		
ETEC		
Norovirus		
Shigella		
Rotavirus		
Adenovirus		
Campylobacter		

CASE DEFINITIONS

Pathogen	Moderate-Severe Clinical Definition	Etiologic confirmation
Malaria	Moderate or severe: Fever >37.5ºC ombined with one or more markers of disease severity including, prostration, respiratory distress, impaired consciousness (Glasgow coma score <11 or Blantyre coma score <3 or AVPU scale P/U), severe anemia, cerebral malaria	Parasitological confirmation by an RDTs, or microscopy, p. falciparum asexual parasitemia (>5000 parasites/mm^3) plus one or more markers of severe disease
Pertussis	Moderate- Severe disease defined as having a 2 weeks of cough plus classic symptoms of pertussis (whooping, paroxysms, apnea, post-tussive vomiting) scored using modified Preziosi Scale (MPS)* [1, 2] . An MPS score of >6 points is severe	Positive PCR or microbiological culture confirmation results in the presence of screening symptoms [1]
RSV	Infants presenting with Lower tract infection (pneumonia or bronchitis) or LRTI associated with hospitalization or LRTI associated with severe hypoxemia or influenza like illness (ILI)** [3]	PCR Positive or antigen test positive and clinical presentation
iNTS	Sufficiently ill to require hospital admission, fever, tachycardia, tachypnea, respiratory distress, and altered consciousness (low Blantyre coma score) and hepatosplenomegaly [4]	Definitive diagnosis based on laboratory confirmation, specifically the isolation of Non-typhoidal Salmonella species by blood culture or cerebrospinal fluid [5]

** Pertussis (MPS scoring) – Paroxysmal cough, inspiratory cough, post-tussive vomiting, cyanosis, apnea, seizures, pneumonia, encephalopathy, hospitalization required*

*** RSV controls – PERCH study included controls with mild upper respiratory symptoms but did not meet case definitions for pneumonia or severe respiratory disease.*

STUDY DESIGN: LIMITATIONS & FUTURE DIRECTIONS

LIMITATIONS OF PROPOSED STUDY DESIGN

01 Exposure to pathogens, particularly rare pathogens, is unknown.

02 Reliance on care-seeking to identify cases can lead to blood samples from diseased children incorrectly used as control samples.

03 Reliance on blood collection to measure immune response misses other types of immunity, such as mucosal immunity.

04 Not all natural infections lead to immunity.

FUTURE DIRECTIONS

01 Extensive literature review to inform **likelihood of each pathogen to elicit systemic immune response** identifiable in blood samples

02 **Longitudinal immune profiling** to track antibody kinetics and immune memory over time in response to infections.

03 **Trade-off analysis of blood collection frequencies:** compare the least conservative (blood collected at disease presentation) versus routine blood collection strategies to determine the optimal balance between practicality and scientific rigor

04 Deeper dive into non-IHME incidence estimates, as well as country-specific incidence estimates to inform cohort sizes

05 Outline **additional questions that would be answerable with the birth cohort dataset** (i.e. co-pathogen [virus-bacteria interaction], disease outcomes, continued AMR surveillance)

QUESTIONS?

START CENTER | STRATEGIC ANALYSIS, RESEARCH & TRAINING CENTER

Department of Global Health | University of Washington

APPENDIX

START CENTER | STRATEGIC ANALYSIS, RESEARCH & TRAINING CENTER

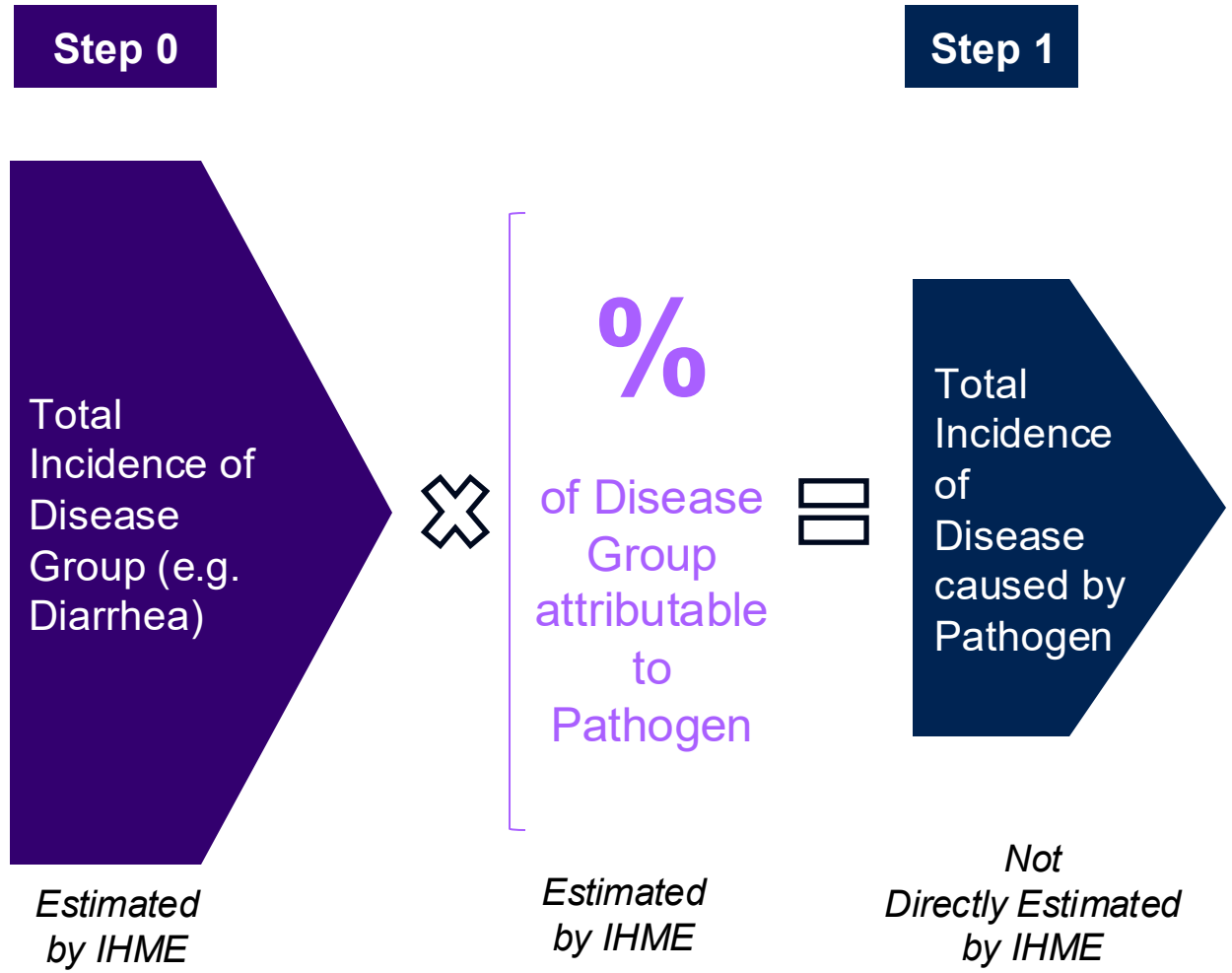
Department of Global Health | University of Washington

BIRTH COHORT SIZE

INCIDENCE RATE CALCULATIONS

Estimating Pathogen-Specific Incidence using GBD Estimates

- For many of our target diseases, the IHME GBD Tool does not directly report pathogen-specific incidence
- To calculate pathogen specific incidence, we multiplied the nonfatal etiology PAF (YLD PAF) by the overall incidence of the broader disease category
- This method can be applied across disease areas (e.g., diarrheal, respiratory, etc.)
- Example: To estimate the incidence of Shigella, multiply the Shigella PAF by the total incidence of diarrheal infections.



BIRTH COHORT SIZE

JOINT SOUTH ASIA (20 cases) & SUB-SAHARAN AFRICA (20 cases)

All incidence rates from IHME

Pathogen	SA Cohort Size	SSA Cohort Size	Joint Cohort Size
Cholera	173,939	41,530	215,469
iNTS	173,521	6,675	180,196
Malaria	67,873	929	68,802
RSV	7,055	10,098	17,153
Pertussis	1,857	1,021	2,878
Cryptosporidium	1,305	478	1,783
ETEC	852	565	1,417
Norovirus	635	554	1,189
Shigella	731	411	1,142
Campylobacter	424	715	1,139
Rotavirus	597	299	896
Adenovirus	295	415	710

BIRTH COHORT SIZE

SHIGELLA BY SEROGROUP/S. FLEXNERI SUBSEROTYPES

Cases = 20 of each Shigella Sero-group or subserotype in South Asia

Shigella Serogroup	% of Total Shigella	Cohort Size
Shigella overall	100%	731
S. flexneri	66%	1,109
S. sonnei	24%	3,083
S. boydii	5.4%	13,543
S. dysenteriae	5.0%	14,752

Shigella Subserotype	% of Total Shigella	Cohort Size
Shigella flexneri overall	66%	1,109
S. flexneri 2a	20%	3,623
S. flexneri 6	11%	6,662
S. flexneri 3a	9.4%	7,794
S. flexneri 1b	7.5%	9,719
S. flexneri 4a	2.9%	25,034

**Proportions of Shigella serogroups and serotypes/subserotypes based on Livio 2014*

iNTS COHORT SIZES FOR HIGH INCIDENCE REGIONS

TOGO, GUINEA, & MALI

Pathogen	Country	Cohort Size
iNTS	Togo	5,490
iNTS	Guinea	5,092
iNTS	Mali	4,754