Measures of Disease in Clinical Epidemiology

Brendan J. Kelly, MD, MS

Infectious Diseases, Epidemiology & Microbiology

University of Pennsylvania

19 November 2024

image: Federica Viazzi "Microscopy image of *P. jirovecii*"

Disclosures

- No conflicts of interest.

Learning Objectives

- Measures of disease occurrence: use population data to describe health and disease.

- Key concepts -- explain the meaning of:

  - data types and terminology
  - prevalence
  - incidence (<u>warning</u>: multiple types)
  - relative risk (RR)
  - odds ratio (OR)

- Implementation: prevalence, incidence, RR, and OR from study data (2x2 tables!).

Why Do I Care? …

- Why use these tools (prevalence, incidence, RR, OR)?

    - inform differential diagnoses & counsel patients (<u>your job</u>)

    - design public health interventions & direct new diagnostics/therapies (society!)

    - understand distributions and determinants of diseases (science!)

... What Do I Need to Know for the Test?

- Why use these tools (prevalence, incidence, RR, OR)?

  - inform differential diagnoses & counsel patients (your job)

  - design public health interventions & direct new diagnostics/therapies (society!)

  - understand distributions and determinants of diseases (science!)

- How to use these tools:

  - precise definitions (e.g., RR vs OR, cumulative incidence vs incidence density)

  - a bit of arithmetic

Data Types
&
Distributions

Data Types
&
Distributions

Case from 1981

Data Types
&
Distributions

Case from 1981

Prevalence

Data Types & Distributions

Case from 1981

Prevalence

Incidence

Data Types & Distributions	Case from 1981	Prevalence
Incidence	Inference from Exposures & Outcomes	

Data Types & Distributions	Case from 1981	Prevalence
Incidence	Inference from Exposures & Outcomes	Relative Risk & Odds Ratios

Data Types

- **Dichotomous**: binary 0 / 1, true / false

  - the focus of most methods in this course (e.g., pneumonia, MI)

- **Continuous**: real values, infinite scale

  - discussed in diagnostics lecture (e.g., age, height)

- **Nominal**: discrete data

  - categories without intrinsic order (e.g., state)

- **Ordinal**: discrete data

  - categories with intrinsic order (e.g., age category)

How to Characterize Continuous Data

- Pair a measure of central tendency with a measure of dispersion:

  - mean and SD (both affected by outliers)

  - median and IQR

- In doing so, account for <u>uncertainty</u> in measures.

Characterize Continuous Data: Central Tendency

- **Mean (μ):**

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$$

- **Median**: middle value, when values ordered / ranked

- **Mode**: most frequently occurring value

(you do **NOT** need to memorize these formulas)

Characterize Continuous Data: Variation/Spread

- **Standard deviation (SD or σ):**

$$\sigma = \sqrt{\frac{1}{N} * \sum_{i=1}^{N}(x_i - \mu)^2}$$

- **Interquartile range (IQR)** depends on ranking the values:

  - the first quartile is the "middle" value of the first half of the ordered set

  - the third quartile is the "middle" value of the second half of the ordered set

  - IQR is the range of values between first and third quartiles

    (you do **NOT** need to memorize these formulas)
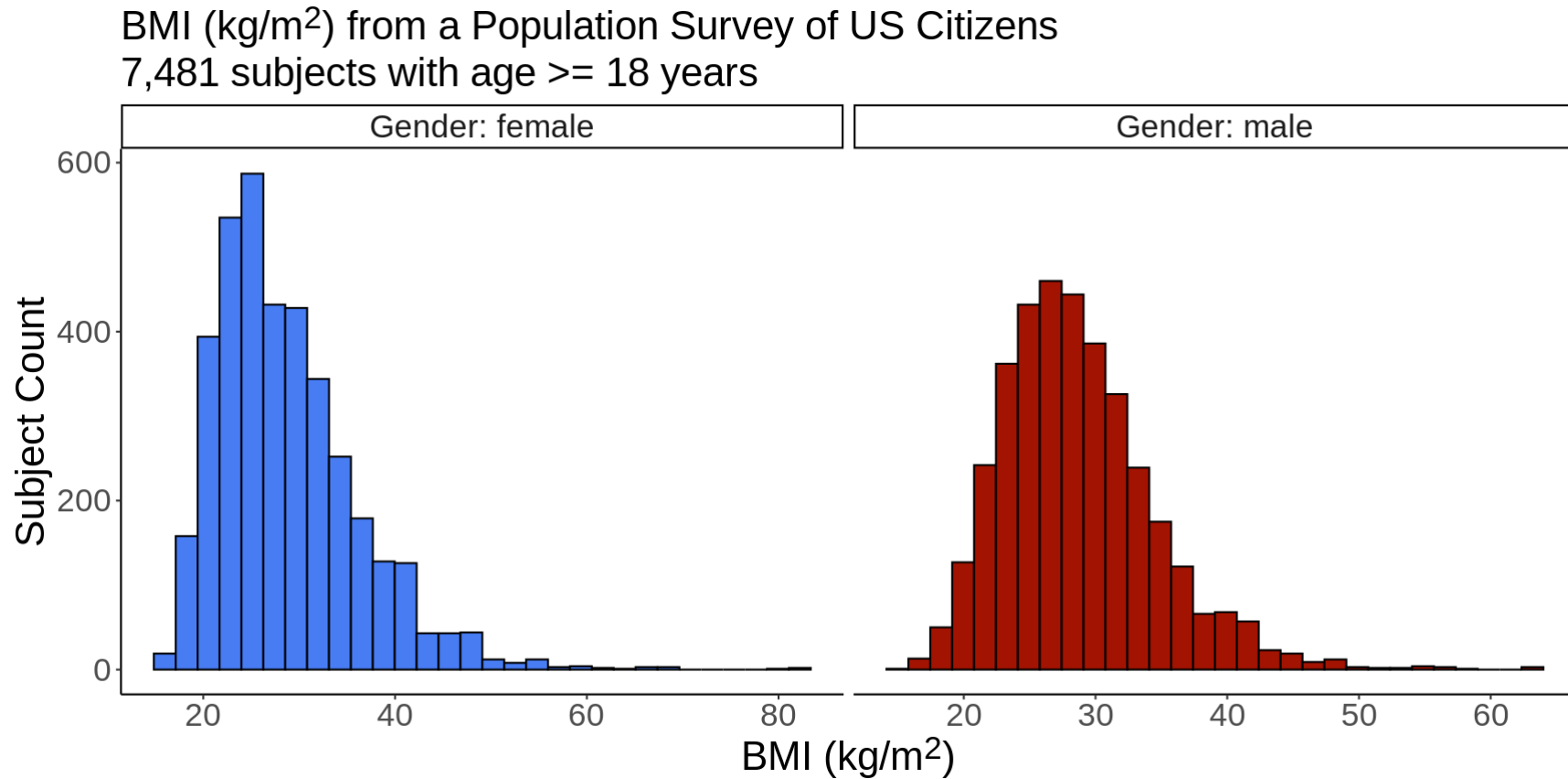
Describing Continuous Data

- "Normally distributed": data well characterized by **mean** and **standard deviation (SD)**.

- "Non-parametric" data have a mean and standard deviation, but these parameters do NOT characterize the data well...

"Normally Distributed" Continuous Data
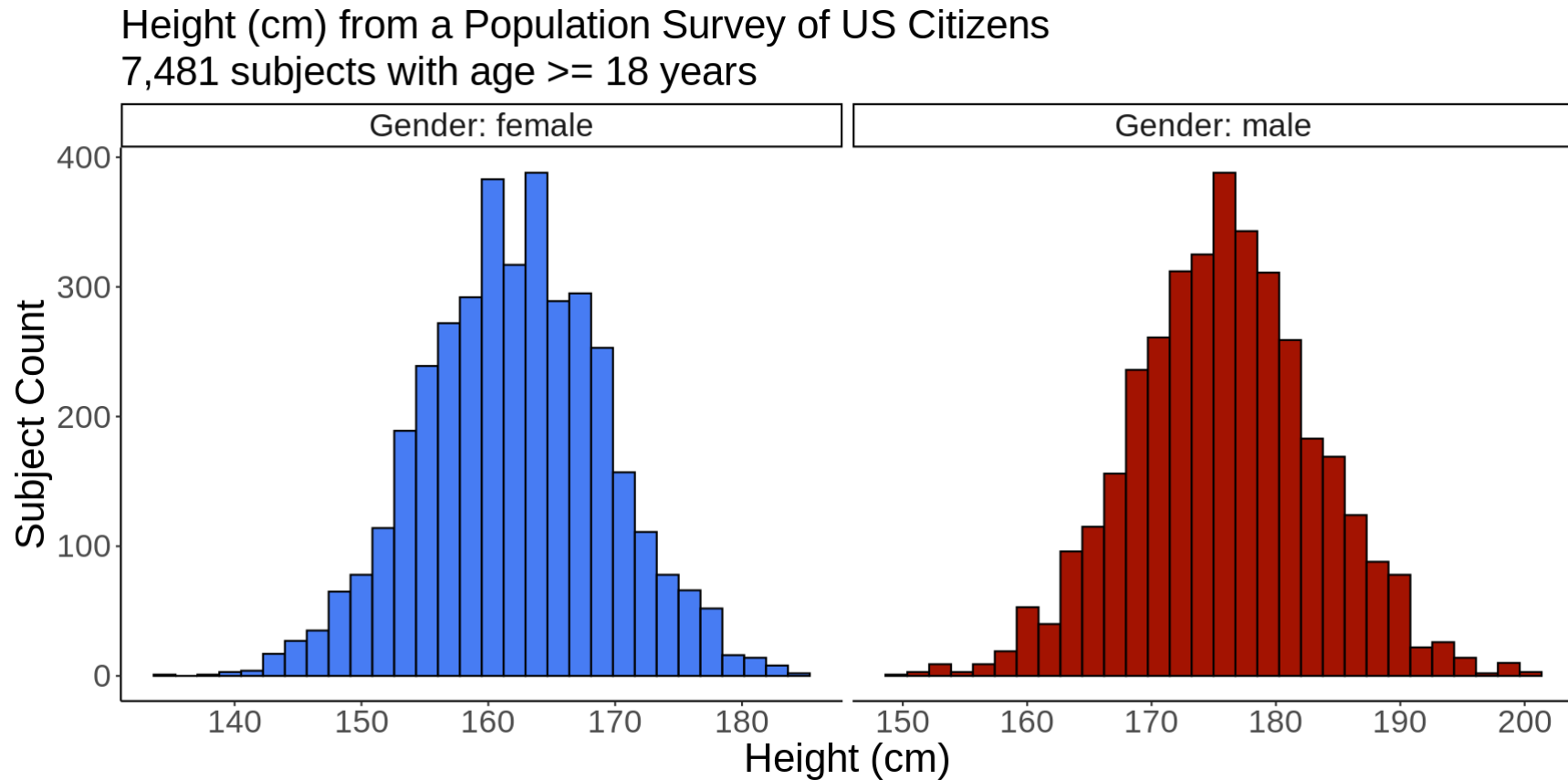
Height (cm) from a Population Survey of US Citizens
7,481 subjects with age >= 18 years

CDC National Health and Nutrition Examination Survey (NHANES)

"Non-Parametric" Continuous Data: Skewed

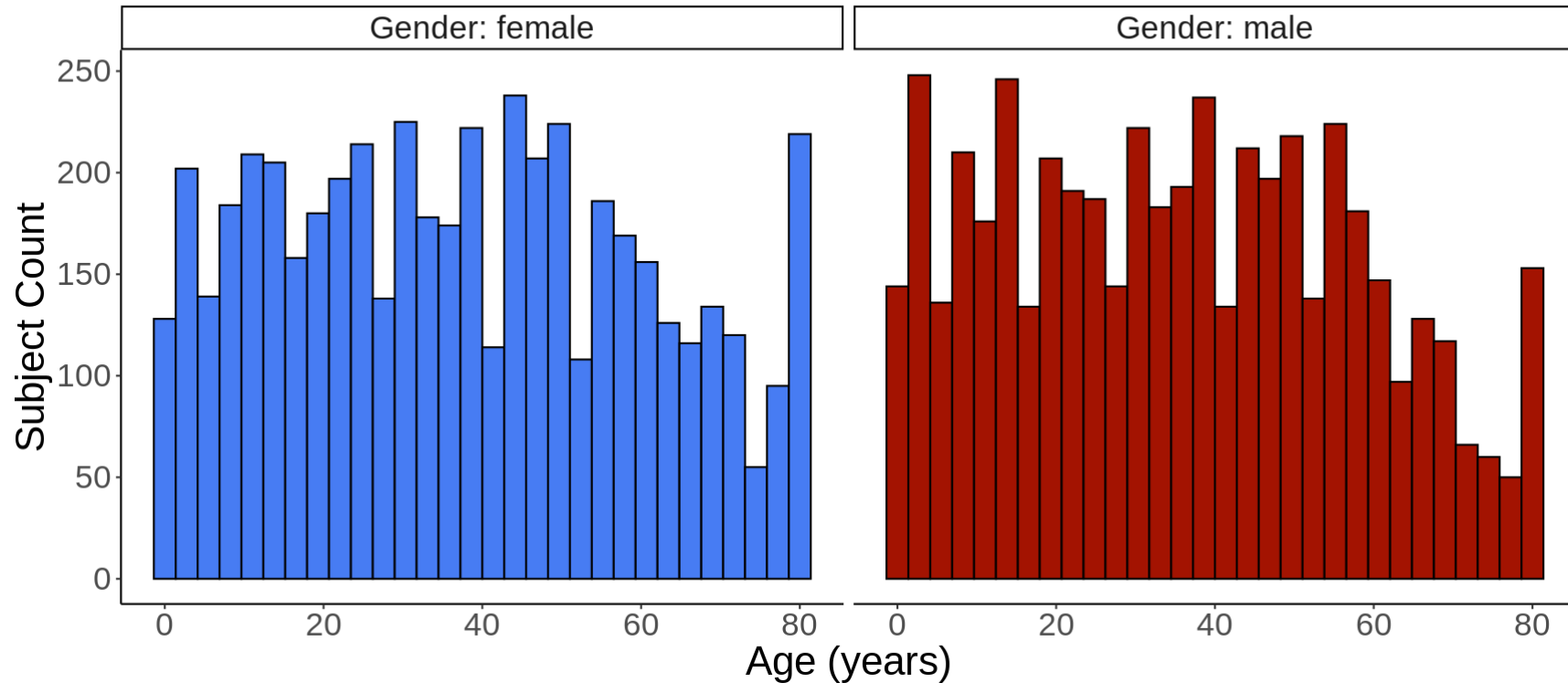BMI (kg/m^2) from a Population Survey of US Citizens
7,481 subjects with age >= 18 years

CDC National Health and Nutrition Examination Survey (NHANES)

"Non-Parametric" Continuous Data: Uniform

Age (years) from a Population Survey of US Citizens

CDC National Health and Nutrition Examination Survey (NHANES)

Describing Continuous Data

- "Normally distributed": data well characterized by **mean** and **standard deviation (SD)**.

- "Non-parametric" data have a mean and standard deviation, but these parameters do NOT characterize the data well:

  - **skewed** data: distribution not symmetric around a central value

  - **uniform** data: distributed with equal probability of each value

- For "non-parametric" data, we prefer to use **median** and **interquartile range (IQR)** to describe the distribution of a continuous variable because these are less affected by extreme values.

- You can always calculate mean and SD from continuous data. But these parameters might not represent the data well.

Dichotomizing Continuous Data

- Using a threshold to transform continuous data into dichotomous data means <mark>losing information about uncertainty</mark>.

- This is on top of the fundamental uncertainty we face with any epidemiologic measure: does the measured population represent the population of interest?

- See the diagnosis lecture: for example, threshold of oxygen saturation (or β-D glucan) to diagnose *Pneumocystis* pneumonia (PCP) means trade-off between sensitivity and specificity.

Dichotomania

- As we discuss prevalence, incidence, RR, and OR, we focus on <u>dichotomous</u> exposures and outcomes.

- Remember -- with dichotomous data:

  - information about uncertainty is lost

  - misclassification is a risk

- But we're doing it anyway 😁 ... why?

- Medicine focuses on dichotomous diagnostic and treatment decisions.

Altman & Royston *BMJ* 2006

What About Regression?

- As we discuss prevalence, incidence, RR, and OR, we focus on <u>dichotomous</u> exposures and outcomes.

- Regression methods account for different data types:

  - linear models: continuous outcomes

  - generalized linear models: dichotomous, nominal, ordinal, etc outcomes

  - can incorporate different exposure data types & multiple exposures

- Regression methods are <u>beyond the scope of this course</u>.

Altman & Royston *BMJ* 2006

Data Types & Distributions	Case from 1981	Prevalence
Incidence	Inference from Exposures & Outcomes	Relative Risk & Odds Ratios

HIV Pandemic: MMWR June 5, 1981

1981

June 5: The U.S. Center for Disease Control (CDC) publishes an article in its *Morbidity and Mortality Weekly Report* (*MMWR*): *Pneumocystis* Pneumonia—Los Angeles. The article describes cases of a rare lung infection, **Pneumocystis carinii pneumonia** (PCP), in five young, white, previously healthy gay men in Los Angeles. Los Angeles immunologist Dr. Michael Gottlieb, CDC's Dr. Wayne Shandera, and their colleagues report that all the men have other unusual infections as well, indicating that their immune systems are not working. Two have already died by the time the report is published and the others will die soon after. This edition of the *MMWR* marks the first official reporting of what will later become known as the AIDS (**Acquired Immunodeficiency Syndrome**) epidemic.

HIV Pandemic: MMWR June 5, 1981

All the above observations suggest the possibility of a <mark>cellular-immune dysfunction related to a common exposure that predisposes individuals to opportunistic infections such as pneumocystosis and candidiasis</mark>. Although the role of CMV infection in the pathogenesis of pneumocystosis remains unknown, the possibility of *P. carinii* infection must be carefully considered in a differential diagnosis for previously healthy homosexual males with dyspnea and pneumonia.

— Gottlieb MS et al <u>MMWR</u> 1981

HIV Pandemic: Stigma & Discrimination

- Socially vulnerable groups often at higher infection risk:

    - HIV, COVID-19

    - how we name infectious diseases has exacerbated stigma

    - *bad epidemiology exacerbates stigma*

- Prejudice is used to gain political power:

    - homophobia in the 1980s

    - anti-Asian racism in the 2020s

Radiolab "The Ashes on the Lawn"

Radiolab

DEC 18, 2020
The Ashes on the Lawn
Image credits: Meg Handler

Transcript

A global pandemic. An afflicted, angry group. A seemingly indifferent government. Reporter Tracie Hunte wanted to understand this moment of pain and confusion by looking back 30 years, and she found a complicated answer to a simple question: When nothing seems to work, how do you make change?

COVID-19 Pandemic: Stigma & Discrimination

- As we proceed in course, think about causal versus proxy variables. For example:

  - Race is not a proxy for biology.

  - Race exerts large effects on health: structural racism & direct / interpersonal racism

- High-risk work, access to care, and social networks are some ways social determinants and racism impact health.

Bilal U et al *Lancet Reg Health Am* 2022;
Chastain DB et al *NEJM* 2020; Lee JH *NEJM* 2021

Racial/ethnic and neighbourhood social vulnerability disparities in COVID-19 testing positivity, hospitalization, and in-hospital mortality in a large hospital system in Pennsylvania: A prospective study of electronic health records

The Lancet Regional Health - Americas
2022;10: 100220
Published online 3 March 2022
https://doi.org/10.1016/j.lana.2022.100220

Usama Bilal,[a,b,*] John B. Jemmott,[c,d] Alina Schnake-Mahl,[b] Kathleen Murphy,[e] and Florence Momplaisir[e,f]

Racial Disproportionality in Covid Clinical Trials

Daniel B. Chastain, Pharm.D., Sharmon P. Osae, Pharm.D., Andrés F. Henao-Martínez, M.D., Carlos Franco-Paredes, M.D., M.P.H., Joeanna S. Chastain, Pharm.D., and Henry N. Young, Ph.D.

N ENGL J MED 383;9   NEJM.ORG   AUGUST 27, 2020

Combating Anti-Asian Sentiment — A Practical Guide for Clinicians

James H. Lee, M.D.   N ENGL J MED 384;25   NEJM.ORG   JUNE 24, 2021

Since March 2020, Asian Americans have experienced an alarming increase in racial discrimination and racially motivated violence.[1] Commentators have attributed this distressing fact to the blame placed on China for causing the Covid-19 pandemic, 12 months, 31% of Asian-American people have reported being subjected to slurs or racist jokes, 26% have feared that someone might threaten or physically attack them, and 58% believe that anti-Asian racism has increased since the beginning of the pan- boost from social media and subsequent attention by national news outlets. This increased exposure has highlighted the vicious nature of these crimes, and Asian Americans are left anxious about the lives of their loved ones and fearful for their own.

Case from 1981... *Pneumocystis* pneumonia?

- 36-year-old man presents with a 4-month history of fever, dyspnea, and cough.

- What is his diagnosis? Does he have *Pneumocystis* pneumonia (PCP)?

- As you collect data (history, vital signs, physical exam, laboratory test values, radiology results), you are making inferences about the diagnosis based on **associations** between your findings and disease states.

Case from 1981... *Pneumocystis* pneumonia?

- Differential diagnosis must be grounded in understanding:

  - distributions of disease: we'll learn about **prevalence** & **incidence**

  - determinants of disease: we'll learn **2x2 tables** to relate exposures and outcomes

- PCP is a fungal pneumonia caused by *Pneumocystis jirovecii* (formerly *Pneumocystis carinii*), which causes dyspnea, hypoxemia, and nonproductive cough, and eventually cystic lung changes.

- In 1981, *Pneumocystis* was known to be a low <u>prevalence</u> disease.

- New data would show an increasing <u>incidence</u>.

Case from 1981… *Pneumocystis* pneumonia?

- Does our patient have *Pneumocystis* pneumonia (PCP)?

- How is the distribution of *Pneumocystis* pneumonia changing?

- What are the determinants of *Pneumocystis* pneumonia?

Note: questions depend on **dichotomous** definition of disease (yes/no *Pneumocystis*)

| Data Types & Distributions | Case from 1981 | Measuring Disease Occurrence |
| Incidence | Inference from Exposures & Outcomes | Relative Risk & Odds Ratios |

Data Types & Distributions	Case from 1981	Prevalence
Incidence	Inference from Exposures & Outcomes	Relative Risk & Odds Ratios

Prevalence

- How common is *Pneumocystis* pneumonia (PCP)?

- Prevalence:

  - number with the disease / number in specified population

  - **point prevalence**: at a specific point in time

  - **period prevalence**: during a given period (e.g., 12-month prevalence)

  - a proportion (unitless, ranges from 0-1)

  - numerator includes all people who have the disease, both new and ongoing cases

  - represents a <u>cross-sectional</u> "snapshot" of the population

Prevalence of *Pneumocystis* pneumonia

- In 1967, CDC became the sole supplier of pentamidine in the United States and began collecting data on cases of PCP:

  - period prevalence published in 1974[*]: 579 cases (194 confirmed) over 3 years
  - what's the denominator?
  - what's the prevalence?

- **Point prevalence** of *Pneumocystis* would be vanishingly small given limited duration of disease.

- From 1967-1974, even the **period prevalence** was very small.

[*] Walzer PD et al *Annals Int Med* 1974

Point Prevalence versus Period Prevalence

- To understand the difference between point and period prevalence, let's imagine a cohort of people living with HIV/AIDS at risk for *Pneumocystis*:

    - 50 high-risk subjects enrolled at the start of a two-year observation period

    - cases of PCP each receive 3 weeks of antibiotic treatment (red on plot)

    - incomplete follow-up (grey on plot)

    - some subjects started on *Pneumocystis* prophylaxis (blue on plot)
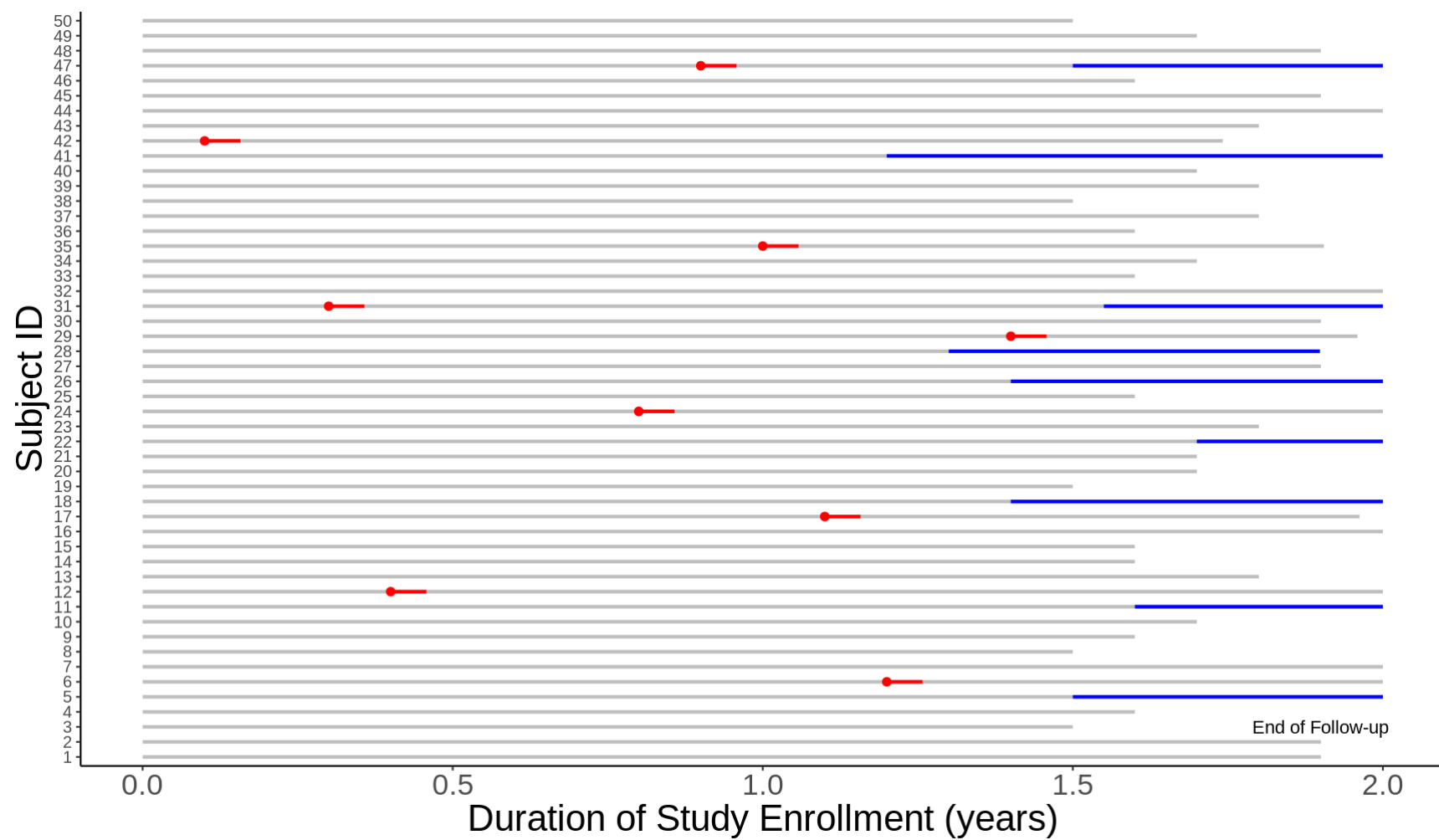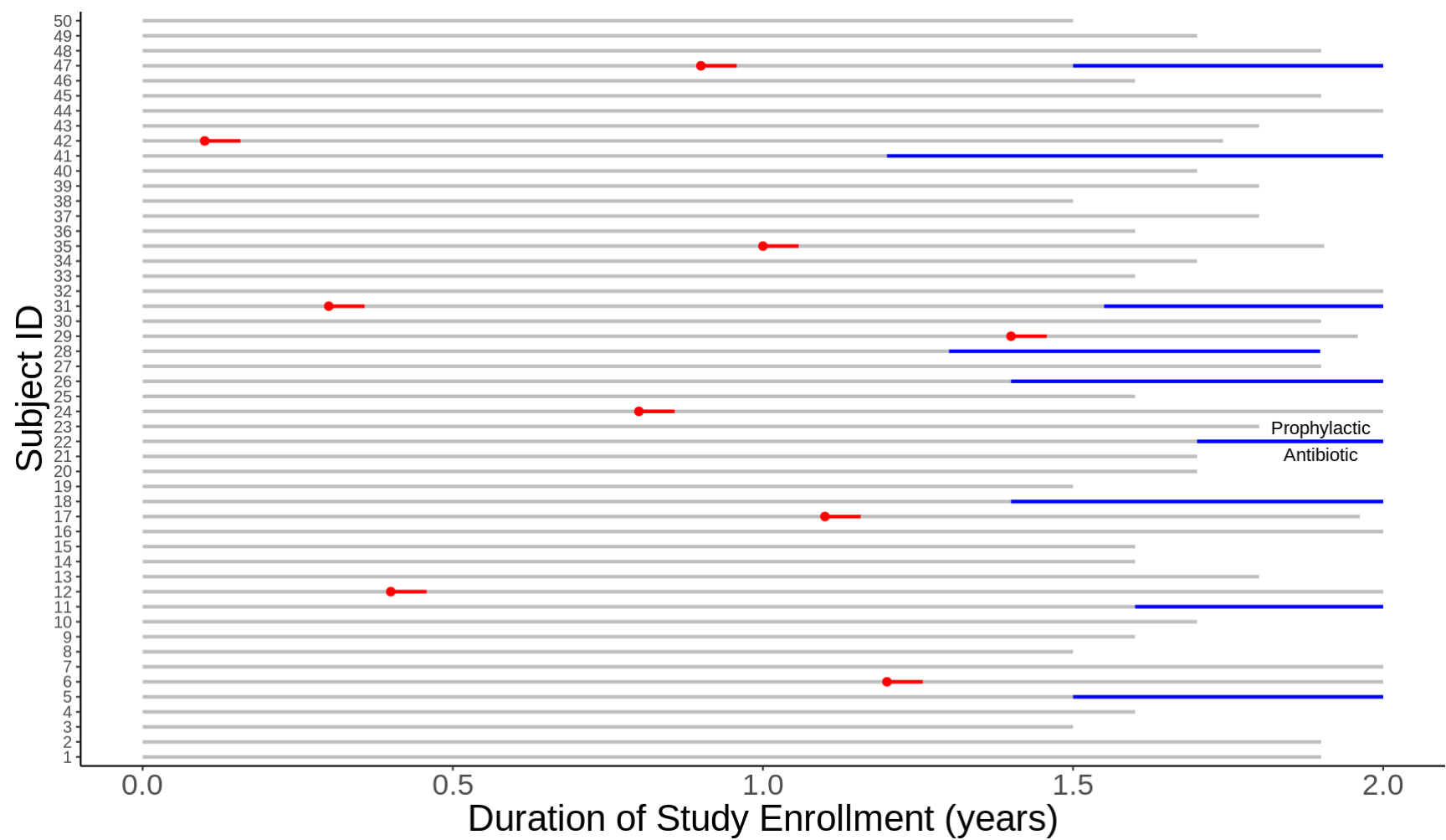
Cohort at Risk for *Pneumocystis*
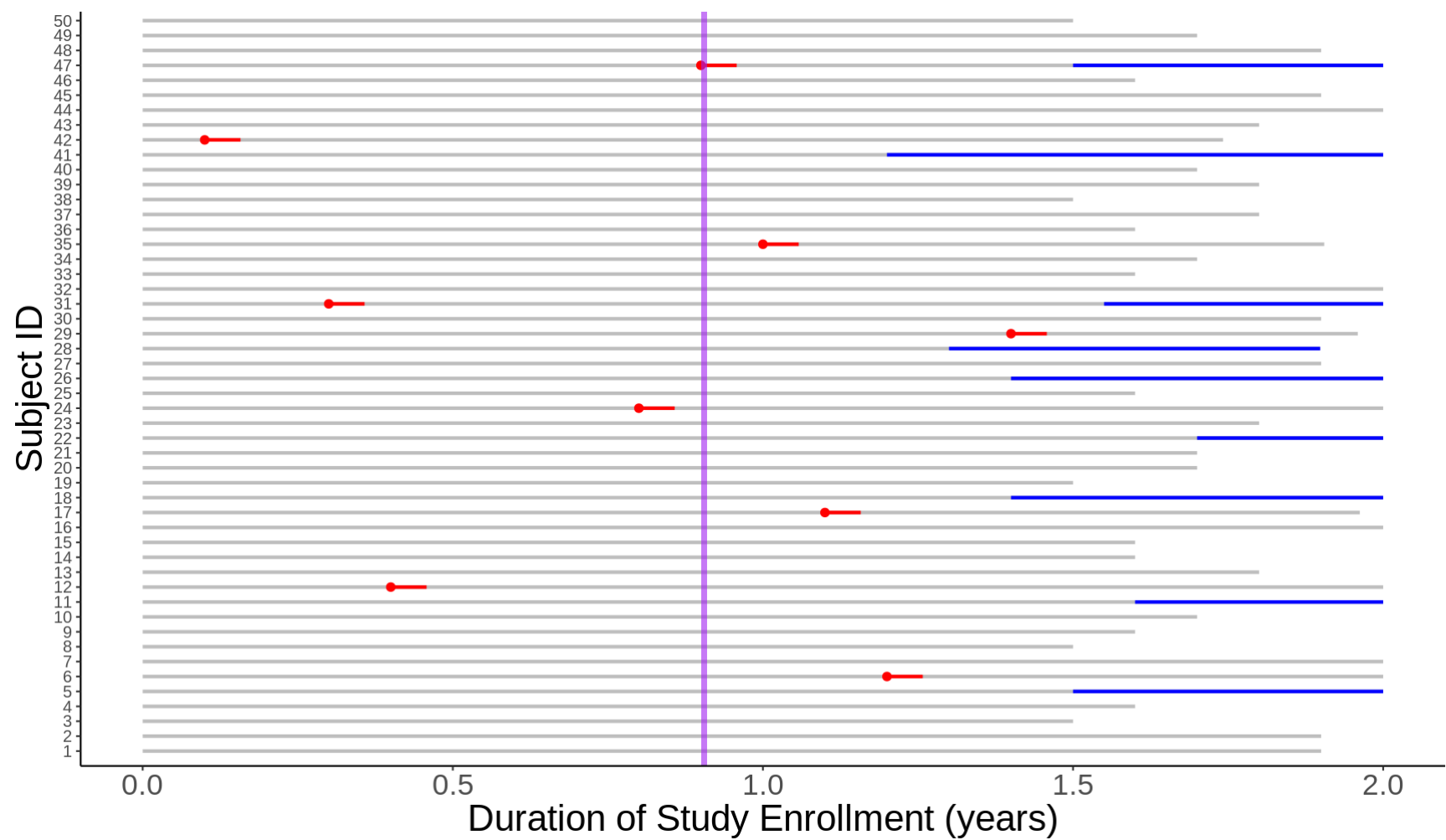
Cohort at Risk for *Pneumocystis*

Cohort at Risk for *Pneumocystis*

End of Follow-up

Subject ID

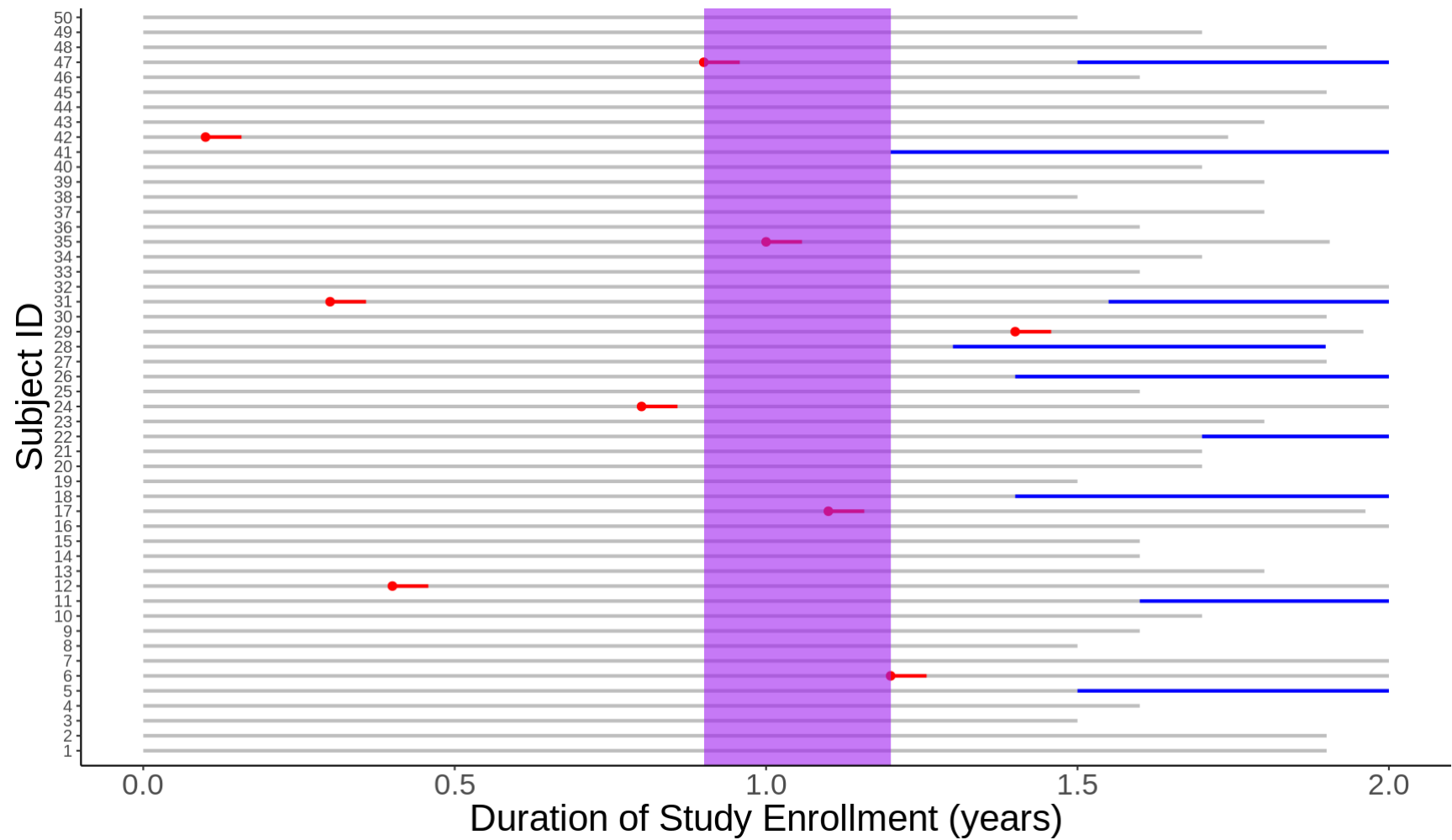Duration of Study Enrollment (years)

40
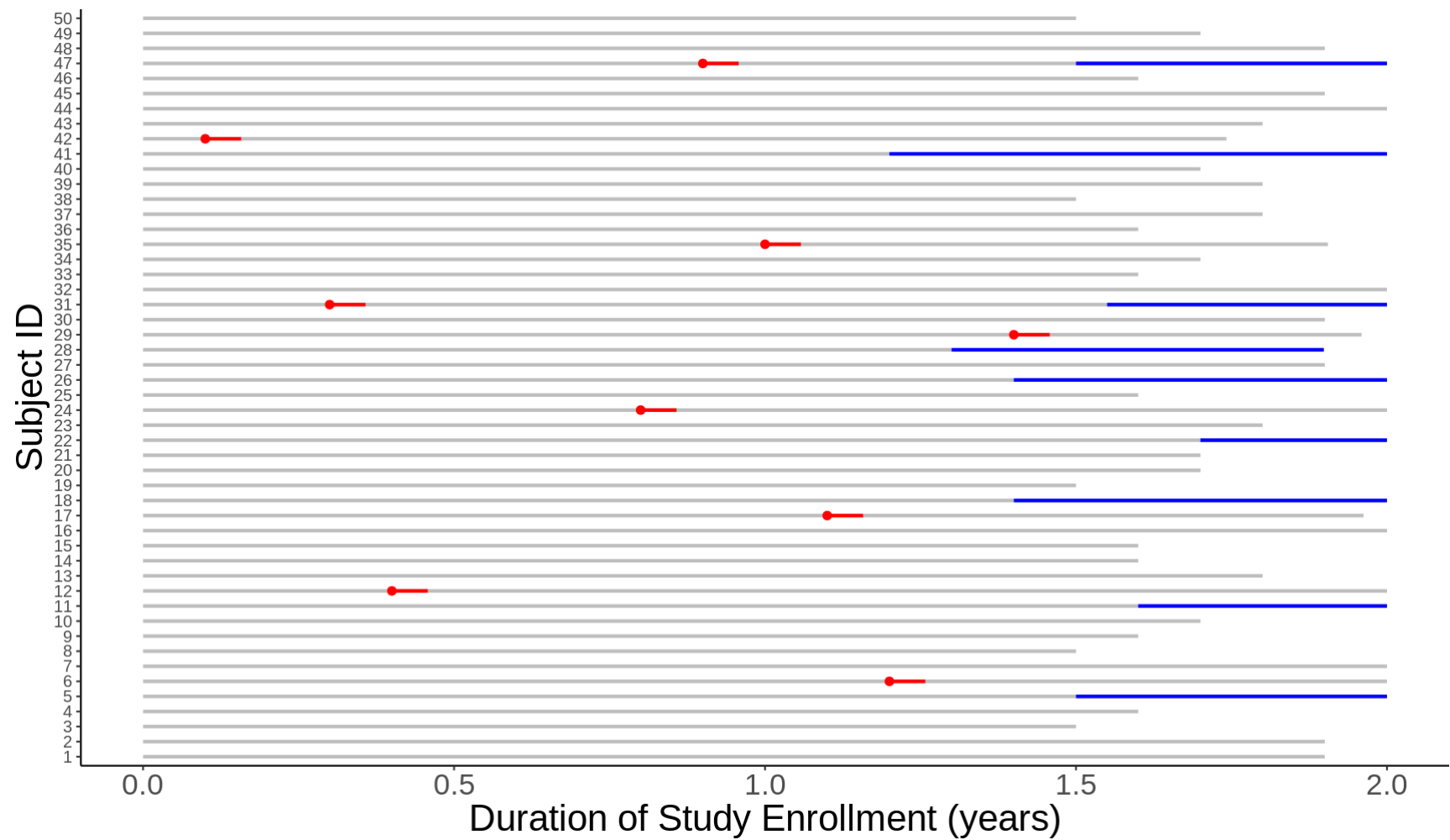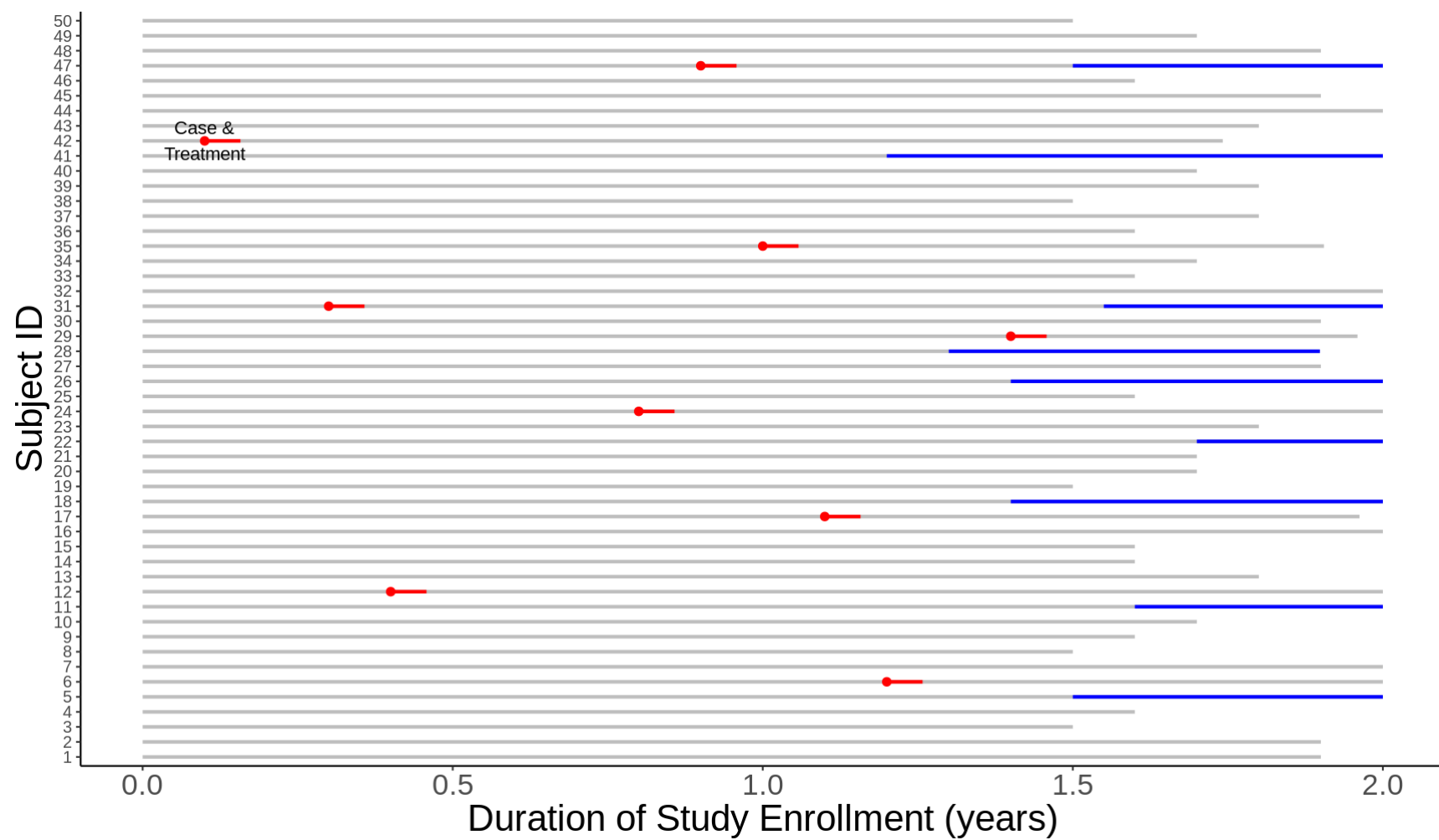
Cohort at Risk for *Pneumocystis*

Cohort at Risk: Point Prevalence

Cohort at Risk: Period Prevalence

Prevalence

- Prevalence is **NOT** the same as risk.

- Prevalence numerator includes all people who have the disease, both new and ongoing cases, so represents a cross-sectional "snapshot" of the population.

- Prevalence does **NOT** estimate the risk of developing the disease because prevalence does **NOT** fully account for time (are the measured cases old cases or new cases?).

- Prevalence can provide a useful snapshot of the burden of disease.

Reflection Question

How can an infection have high prevalence if it occurs infrequently?

- (A) the infection is rapidly fatal

- (B) the infection rapidly resolves

- (C) a few children get the infection every year, but the infection persists for the rest of their lives

- (D) the infection results in lifelong protective immunity

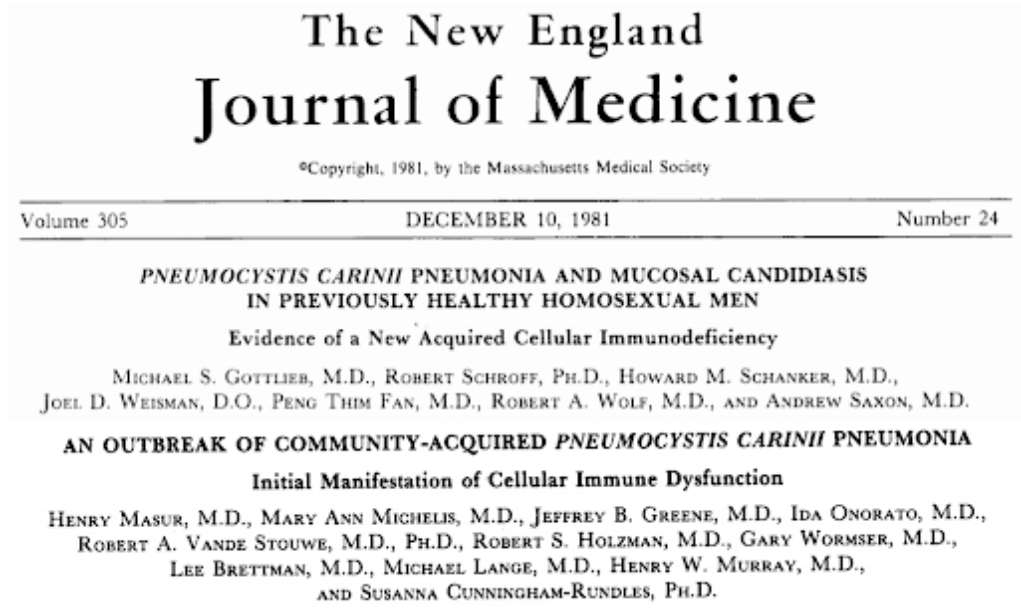| Data Types & Distributions | Case from 1981 | Prevalence |
| Measuring Disease Occurrence | Inference from Exposures & Outcomes | Relative Risk & Odds Ratios |

| Data Types & Distributions | Case from 1981 | Prevalence |
| Incidence | Inference from Exposures & Outcomes | Relative Risk & Odds Ratios |

Case from 1981... *Pneumocystis* pneumonia?

The New England Journal of Medicine

©Copyright, 1981, by the Massachusetts Medical Society

Volume 305   DECEMBER 10, 1981   Number 24

PNEUMOCYSTIS CARINII PNEUMONIA AND MUCOSAL CANDIDIASIS IN PREVIOUSLY HEALTHY HOMOSEXUAL MEN

Evidence of a New Acquired Cellular Immunodeficiency

MICHAEL S. GOTTLIEB, M.D., ROBERT SCHROFF, PH.D., HOWARD M. SCHANKER, M.D., JOEL D. WEISMAN, D.O., PENG THIM FAN, M.D., ROBERT A. WOLF, M.D., AND ANDREW SAXON, M.D.

AN OUTBREAK OF COMMUNITY-ACQUIRED *PNEUMOCYSTIS CARINII* PNEUMONIA

Initial Manifestation of Cellular Immune Dysfunction

HENRY MASUR, M.D., MARY ANN MICHELIS, M.D., JEFFREY B. GREENE, M.D., IDA ONORATO, M.D., ROBERT A. VANDE STOUWE, M.D., PH.D., ROBERT S. HOLZMAN, M.D., GARY WORMSER, M.D., LEE BRETTMAN, M.D., MICHAEL LANGE, M.D., HENRY W. MURRAY, M.D., AND SUSANNA CUNNINGHAM-RUNDLES, PH.D.

- new *Pneumocystis* cases (NYC[1]): 13 cases over 21 months

- new *Pneumocystis* cases (LA[2]): 5 cases over 7 months

- concurrent opportunistic infections:

  - CMV
  - *Candida*
  - Kaposi's sarcoma

[1,2] Masur H et al *NEJM* 1981; Gottlieb MS et al *NEJM* 1981

Incidence

- *Pneumocystis* pneumonia (PCP) is occurring more frequently...

- Incidence: occurrence of new cases over a given period of time.

- **cumulative incidence**:

$$cumulative\ incidence = \frac{new\ cases}{persons\ at\ risk\ at\ start\ time\ interval}$$

- **incidence density**: (more precise)

$$incidence\ density = \frac{new\ cases}{person\ time\ at\ risk}$$

Cumulative Incidence

- Cumulative incidence:

    - must specify population consisting of at-risk individuals

    - must specify a time period of observation

    - <u>numerator</u> = all new cases during a specified time period

    - <u>denominator</u> = all individuals at risk in the specified population at the start of the specified time period (does NOT account for deaths due to other causes)

    - ranges from 0 to 1 (a.k.a., "incidence proportion")

    - like prevalence, is a proportion and therefore has no units (but only makes sense if you specify the time period of observation, e.g., % per year)

Incidence Density

- Incidence density:

  - in a specified population consisting of at risk individuals over a specified period of observation, more precisely quantifies the person-time at risk

  - <u>numerator</u> = all new cases during a specified time period

  - <u>denominator</u> = the sum, over all individuals in the population, of time at risk until the event of interest, death, loss to follow-up, the end of the study, or when they are no longer at risk for whatever reason

  - not a proportion; range depends on the units of person-time (0 to infinity)

 - <mark>accounts for death from other causes!</mark>

Incidence: Which Denominator?

- To understand the difference between cumulative incidence and incidence density, let's return to our imagined study of people living with HIV/AIDS at risk for *Pneumocystis*:

  - 50 high-risk subjects enrolled at the start of a two-year observation period

  - cases of PCP each receive 3 weeks of antibiotic treatment (red on plot)

  - incomplete follow-up (grey on plot)

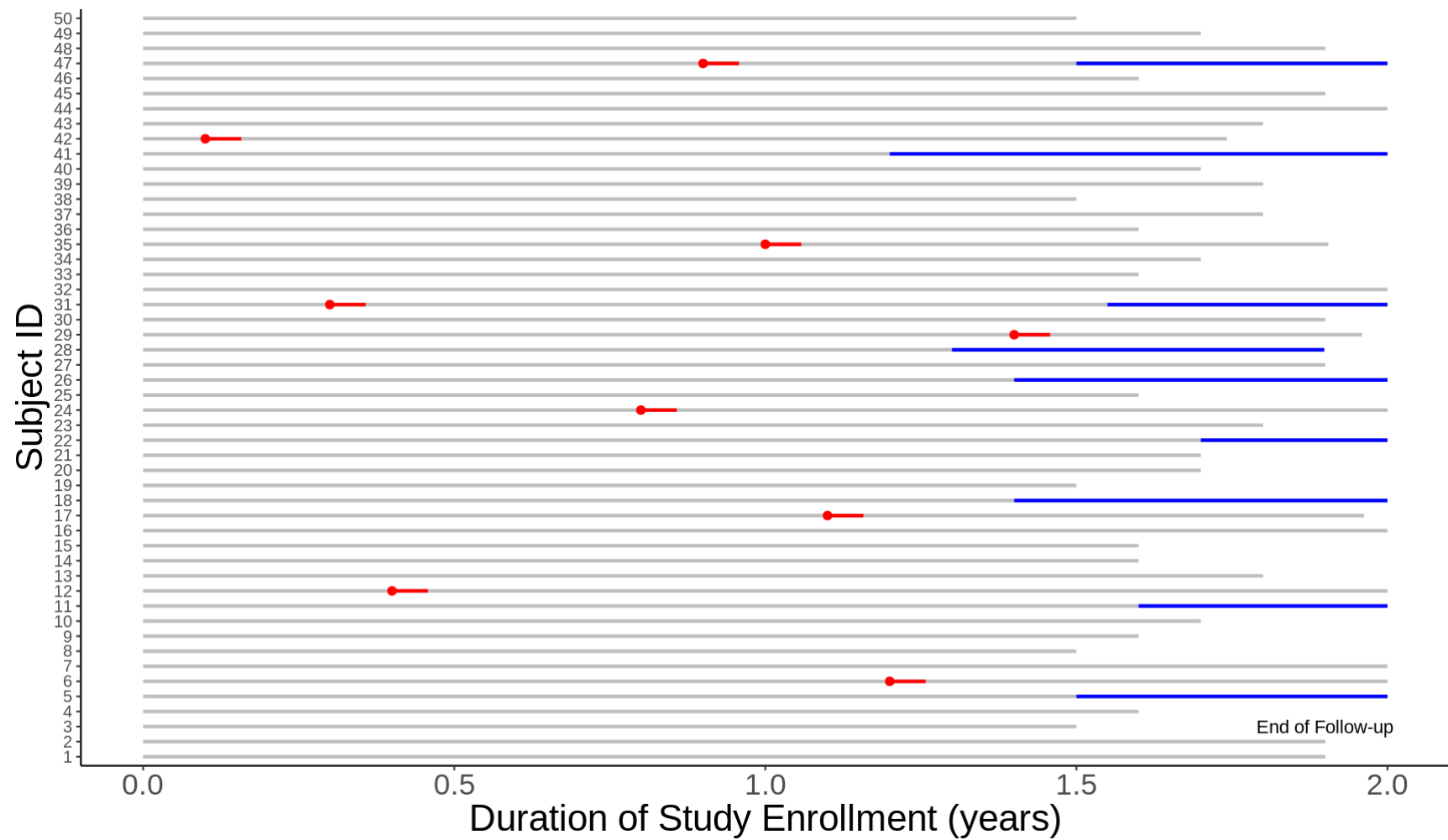 - some subjects started on *Pneumocystis* prophylaxis (blue on plot)
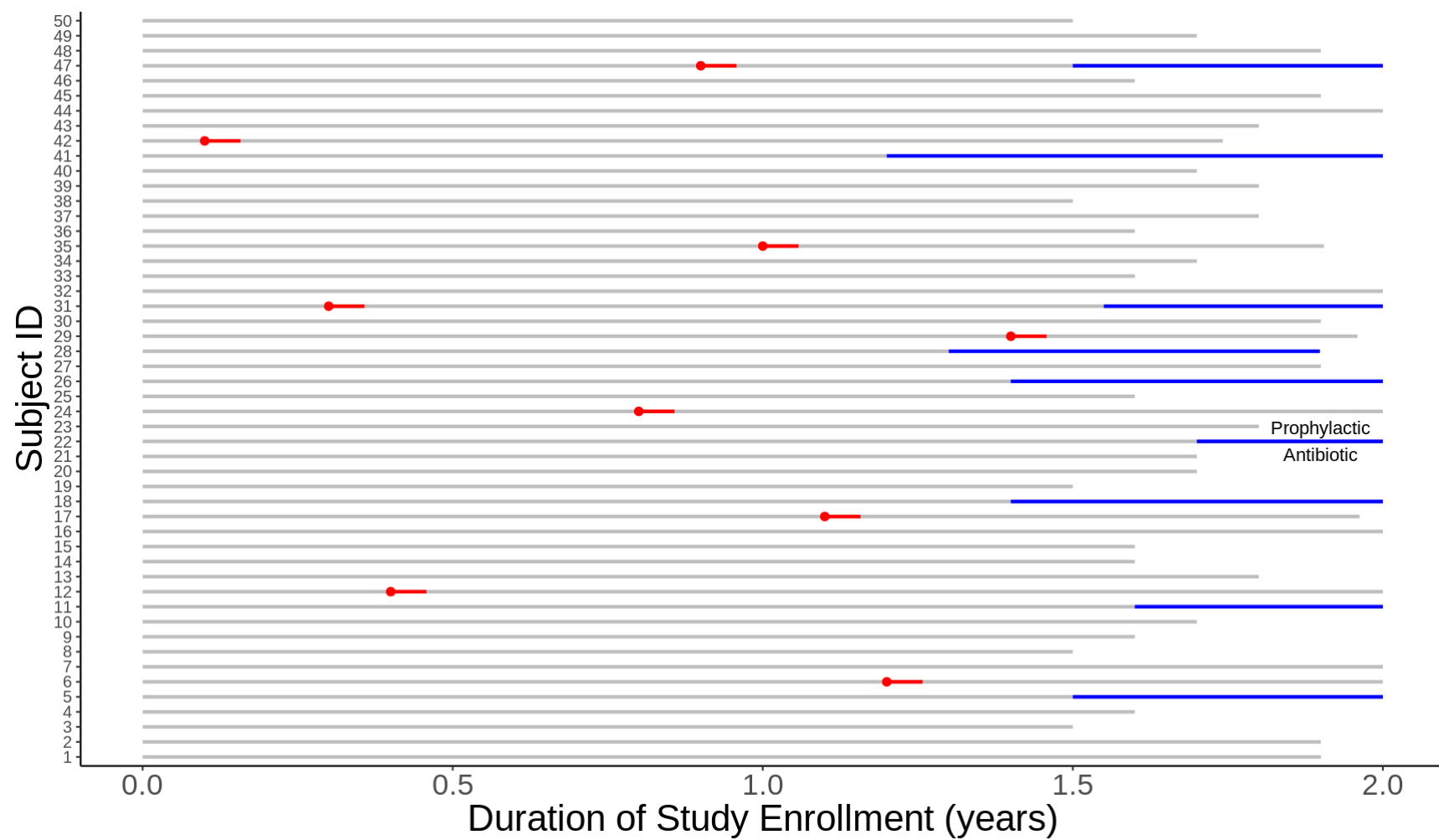
Cohort at Risk for *Pneumocystis*
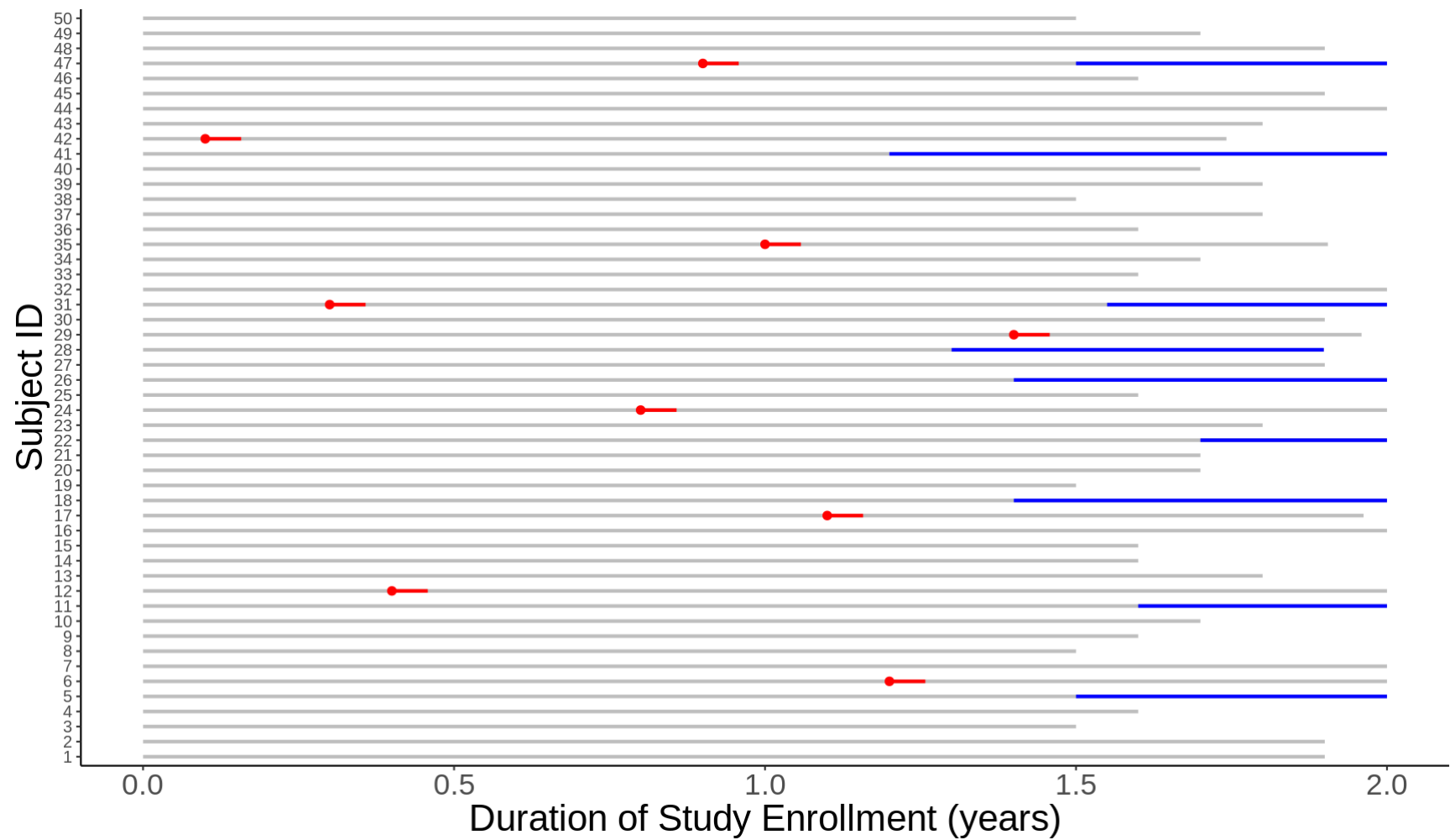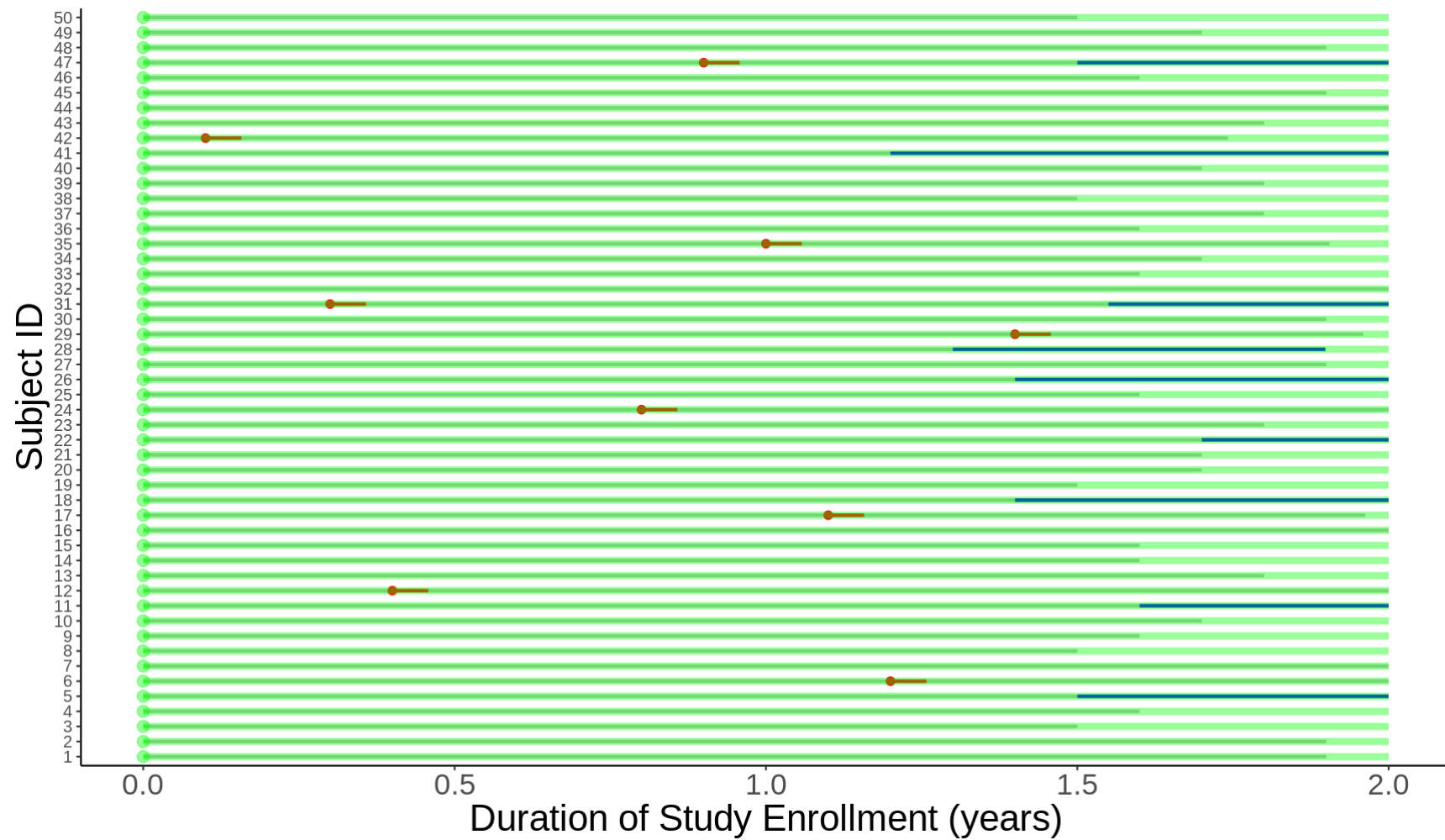
Cohort at Risk for *Pneumocystis*
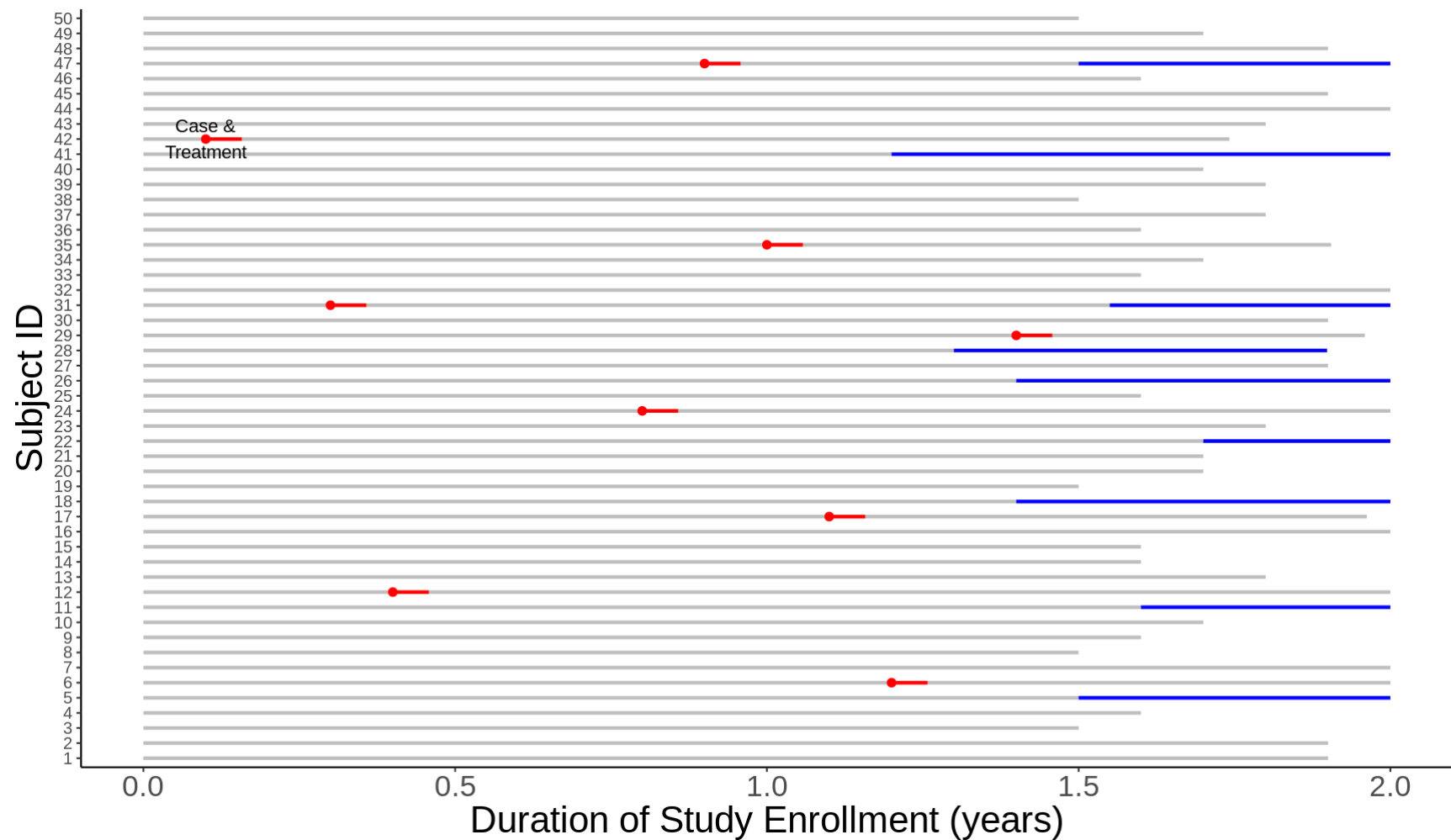
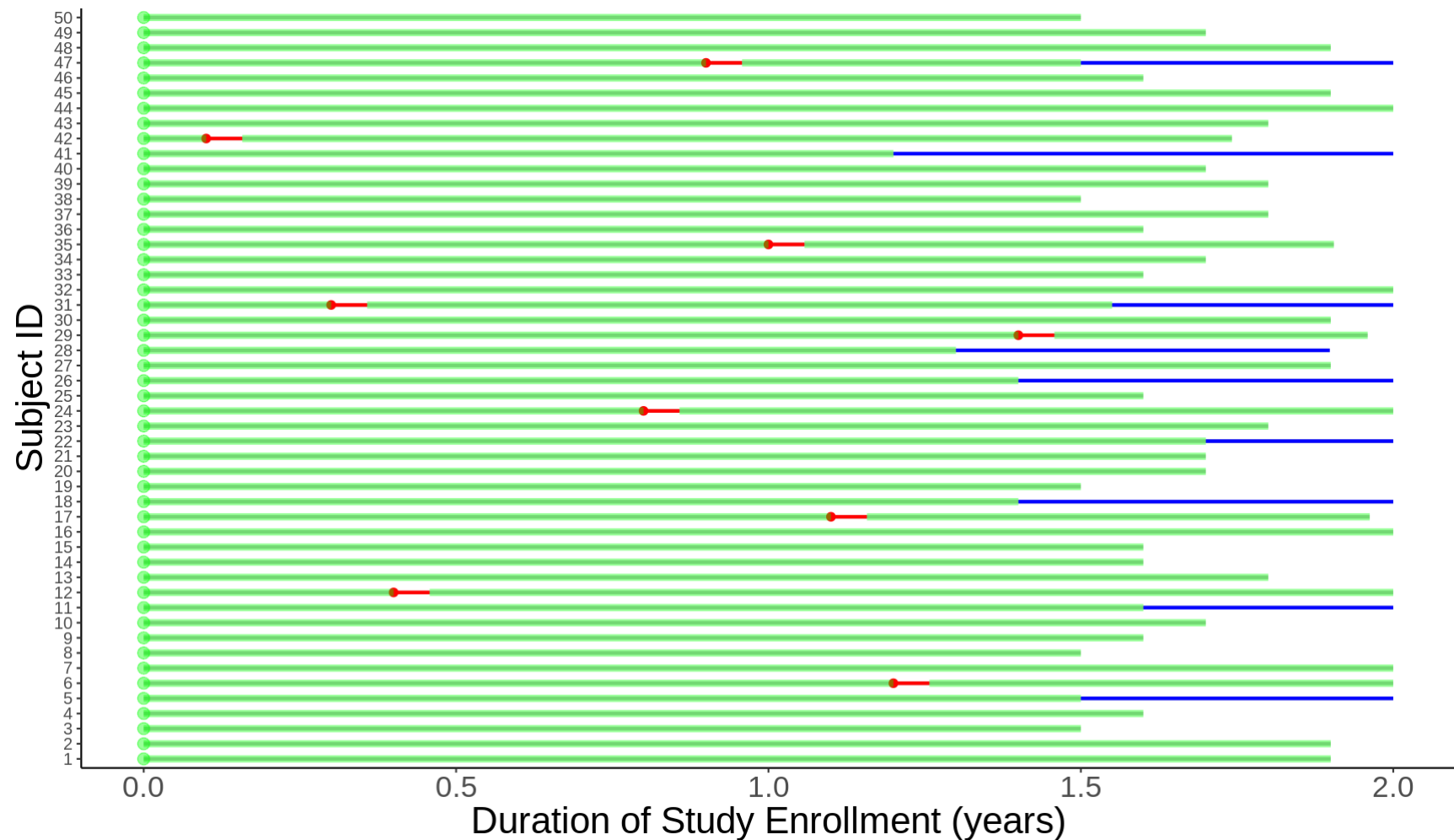# Cohort at Risk for *Pneumocystis*

Cohort at Risk for *Pneumocystis*

Cohort at Risk: Cumulative Incidence (at risk at start)

Cohort at Risk: Incidence Density (person-time at risk)

Incidence: Which Denominator?

- Cumulative incidence of *Pneumocystis* versus incidence density:

  - if the end of the grey line is death / loss to follow-up, how does incidence density compare to annual cumulative incidence?

  - if you don't count time on prophylaxis or treatment antibiotics as "time at risk", how does the incidence density compare to the annual cumulative incidence?

Notes on Population at Risk

- In a population, individuals are at risk of disease if they:

  - (1) do not have the disease at baseline

  - (2) are capable of developing the disease (e.g., have the organ of interest; have not been successfully immunized against the disease; haven't developed lifelong immunity)

- The difference between cumulative incidence and incidence density is that the latter attempts a more precise quantification of population at risk -- it's harder to evaluate, but more informative if you can.

Notes on Person-Time

- To improve precision of risk estimate (<u>incidence density</u>), we measure:

  - population at risk of outcome

  - actual time at risk (during which outcome could occur)

- **Person-time**:

  - denominator for incidence density

  - the product of population at risk and time at risk (e.g., patient-months)

Beware the Phrase "Incidence Rate"!

- "Incidence rate" is used to mean two different things:

  - number new cases / number persons at risk @ start (short) time interval (e.g., "annual incidence rate" to mean cumulative incidence over one year)

  - number new cases / person-time at risk (i.e., incidence density, the precise rate)

Reflection Question

Your patient with HIV is considering starting prophylactic antibiotics for PCP. You have PCP prevalence, cumulative incidence, and incidence density data available. Which data provide the most precise information on the patient's risk of PCP off of prophylaxis?

- (A) prevalence

- (B) cumulative incidence

- (C) incidence density

Can You Tell Prevalence from Incidence?

Incidence

Prevalence

Population Size

Death or Recovery
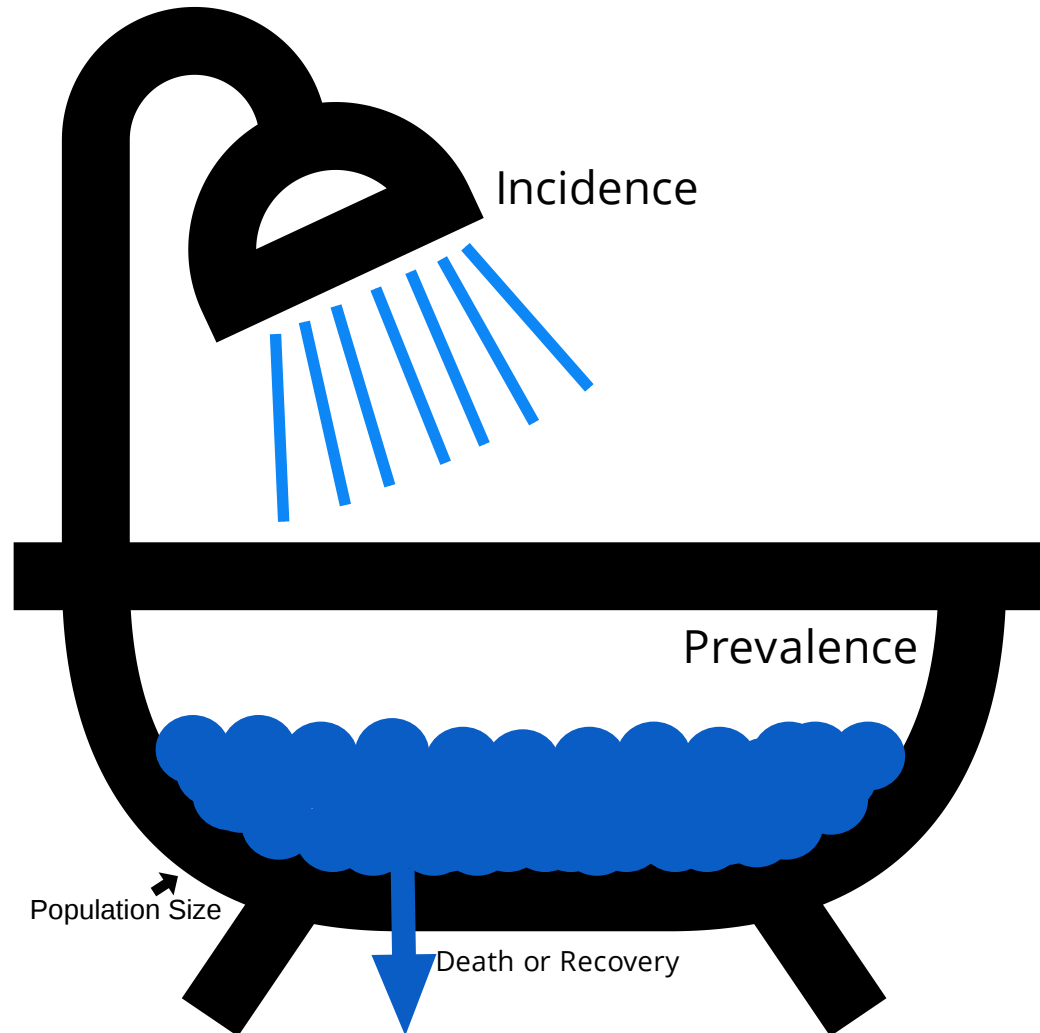
Can You Tell Prevalence from Incidence?

- HIV in Rakai, Uganda 1994-2003*:

    - intensive "ABC" intervention (Abstinence, Be faithful, Condoms)

    - prevalence declined…

    - incidence remained constant at 1.5% per year!

    - what happened?

[*] Wawer M et al *CROI* 2005; Roehr B *BMJ* 2005

Data Types & Distributions	Case from 1981	Prevalence
Incidence	**What determines risk?**	Relative Risk & Odds Ratios

| Data Types & Distributions	Case from 1981	Prevalence
Incidence	**Inference from Exposures & Outcomes**	Relative Risk & Odds Ratios

Basics of Study Types

- We want to understand the ==relationship between risk factors (exposures) and disease (outcomes)==. Move beyond risk in whole population -- what's risk conditional on particular risk factors?

- To calculate incidence need to know how many are in a population:

    - randomized trials: pick the population, randomize, control the treatment, and measure the outcome

    - cohort studies: pick the population, divide into preselected exposure (treatment or risk factor) groups, and measure the outcome

- But do NOT know this in case control studies: pick the cases and control groups, then measure rates of exposure (do NOT know size of population at risk).
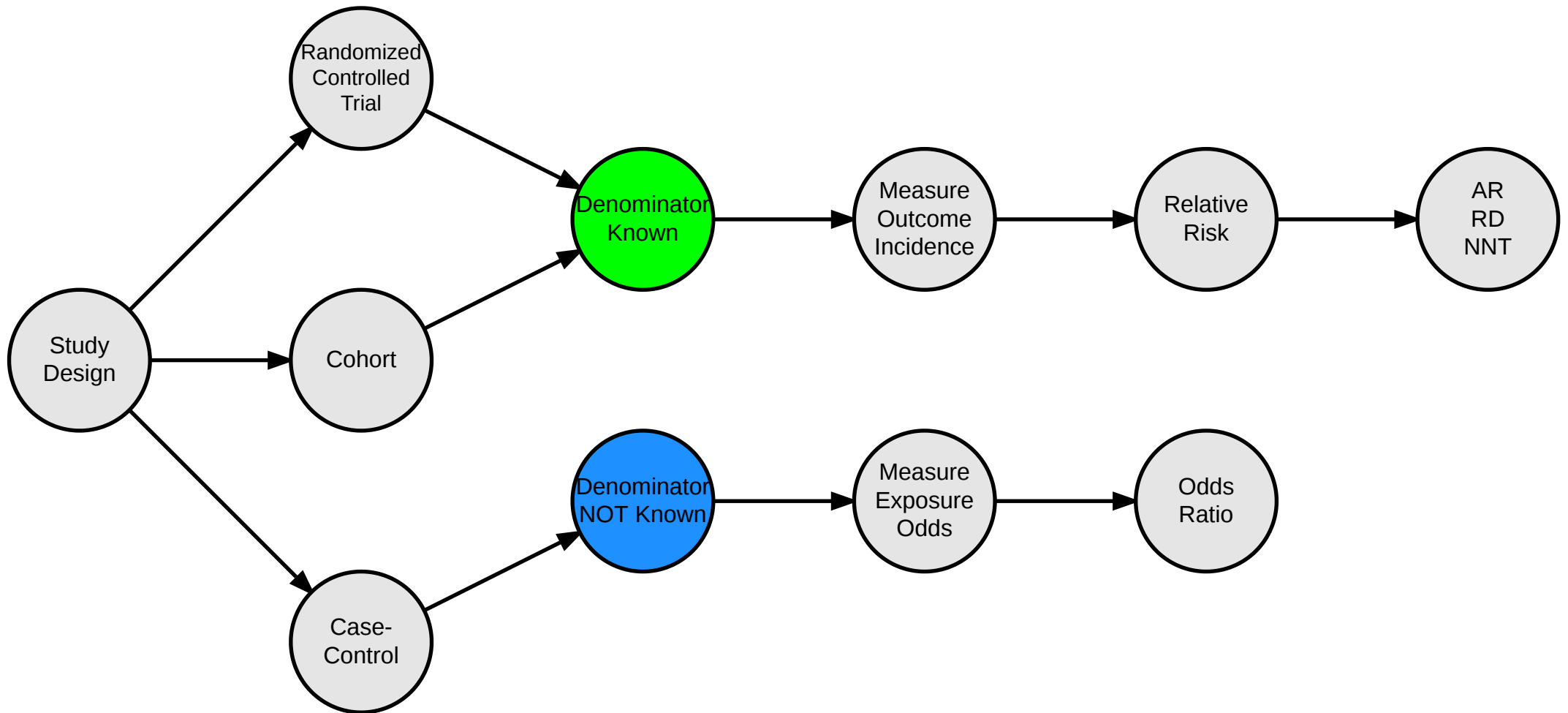
2x2 Table

- Dichotomous exposures and outcomes.

- Examine relationships between exposures and outcomes.

- Goal: inference about larger world.

Outcome

	+	-
Exposure +	A	B
Exposure -	C	D

2x2 Table

- Dichotomous exposures and outcomes.

- Examine relationships between exposures and outcomes.

- Goal: inference about larger world.

2x2 Table: PICO

- Relationship to your evidence-based medicine (EBM) course:

    - PICO: population, intervention, control, outcome

    - intervention/control are exposure groups; outcome is dichotomous for 2x2 table data

- <u>Goal</u>: inference about larger world.

Pneumocystis

	+	-
AIDS +	10	10
AIDS -	5	15

2x2 Table

- Dichotomous exposures and outcomes.

- Examine relationships between exposures and outcomes.

- Goal: inference about larger world.

- Is study RCT/cohort or case-control?

- Can always calculate a relative risk (RR) from 2x2 table but only appropriate for RCT/cohort

- Can always calculate an odds ratio (OR) from 2x2 table but only appropriate for case-control study
  (can do better with RR if RCT/cohort)

Pneumocystis

	+	-
AIDS +	10	10
AIDS -	5	15

Data Types & Distributions	Case from 1981	Prevalence
Incidence	Inference from Exposures & Outcomes	**Aggregate vs Conditional Risk**

Data Types & Distributions	Case from 1981	Prevalence
Incidence	Inference from Exposures & Outcomes	**Relative Risk & Odds Ratios**

2x2 Table: Calculate Relative Risk for Cohort/Trial

- Relative Risk (RR): risk (incidence) exposed / risk (incidence) unexposed:

$$RR = \frac{\frac{A}{(A+B)}}{\frac{C}{(C+D)}}$$

- Risk difference (RD): risk (incidence) exposed - risk (incidence) unexposed:

$$RD = \frac{A}{(A+B)} - \frac{C}{(C+D)}$$

Outcome

+   -

Exposure

+   A   B

-   C   D

PCP ~ AIDS: Relative Risk (RR)

- Imagine a cohort study examining incidence (risk) for *Pneumocystis* among patients with or without an Acquired Immunodeficiency Syndrome (AIDS):

- Relative Risk (RR):

$$RR = \frac{\frac{10}{(10+10)}}{\frac{5}{(5+15)}} = \frac{0.5}{0.25} = 2$$

- Risk difference (RD):

$$RD = \frac{10}{(10+10)} - \frac{5}{(5+15)} = 0.25$$

77

2x2 Table: Calculate NNT for Cohort/Trial

- Risk difference (RD): risk (incidence) exposed - risk (incidence) unexposed:

$$RD = \frac{A}{(A+B)} - \frac{C}{(C+D)}$$

- Number needed to treat (NNT): given RD between exposures, how many exposure switches needed to change one outcome:

$$NNT = \frac{1}{RD} = \frac{1}{\frac{A}{(A+B)} - \frac{C}{(C+D)}}$$

Outcome

	+	-
Exposure +	A	B
-	C	D

78

PCP ~ AIDS: Number Needed to Treat

- Cohort study examining incidence (risk) for *Pneumocystis* among patients with or without AIDS:

- Risk difference (RD):

$$RD = \frac{10}{(10 + 10)} - \frac{5}{(5 + 15)} = 0.25$$

- Number needed to treat (NNT): given RD between exposures, how many exposure switches needed to change one outcome:

$$NNT = \frac{1}{RD} = \frac{1}{\frac{10}{(10+10)} - \frac{5}{(5+15)}} = \frac{1}{0.25} = 4$$

Pneumocystis

	+	-
AIDS +	10	10
AIDS -	5	15

AR vs RD vs NNT

- **absolute risk (AR)**: risk of developing disease over a period of time (<mark>incidence!</mark>)

  - if 1 in 10 chance of developing skin cancer in your lifetime, AR = 10%

- **risk difference (RD)**: difference in risk between treatment/exposure and control

  - 3 in 10 cured with treatment vs 2 in 10 with control, RD = 3/10 − 2/10 = 10%

- **number needed to treat (NNT)**: number treated for one person to benefit

  - NNT = 1/RD
  - from RD numbers above, NNT = 1/RD = 1/0.1 = 10 (treat 10 people to cure 1 more)

2x2 Table: Calculate Odds Ratio for Case-Control Study

- Denominators are deceptive in case-control study because determined by investigator.

- Impossible to measure true risk/incidence. Instead, measure exposure odds in cases & controls.

- Odds Ratio (OR): odds exposure in cases / odds exposure in controls:

$$OR = \frac{\frac{A}{C}}{\frac{B}{D}} = \frac{A * D}{B * C}$$

Outcome

+   -

Exposure

+   A   B

-   C   D

PCP Cases vs Controls: Odds Ratio (OR)

- Imagine a different study: 100 *Pneumocystis* cases and 100 controls; measure whether 'AIDS' (exposure) differs between groups chosen by investigator.

- Odds Ratio (OR): odds exposure in cases / odds exposure in controls:

$$OR = \frac{\frac{90}{10}}{\frac{50}{50}} = \frac{90 * 50}{50 * 10} = 9$$

Pneumocystis

	+	-
AIDS +	90	50
AIDS -	10	50

82

Utility of Odds Ratio (OR) & Case-Control Study Design

- Because ratio of cases and controls determined by investigator, not valid to calculate incidence of outcome (AR or RR, RD, NNT), but if outcome incidence is low, then A + B ~ B & C + D ~ D:

$$RR = \frac{\frac{A}{(A+B)}}{\frac{C}{(C+D)}} \sim \frac{\frac{A}{B}}{\frac{C}{D}} = \frac{A*D}{B*C} = \frac{\frac{A}{C}}{\frac{B}{D}} = OR$$

- OR approximates RR if outcome occurs infrequently (<15%). ==Though OR is fundamentally different from RR, we use it as an approximation of RR==. If outcome is more common, OR & RR differ increasingly.

Outcome

	+	-
Exposure +	A	B
Exposure -	C	D

Measures of Disease in Clinical Epidemiology

- Distribution of data determines how we describe them: mean + SD vs median + IQR.

- Prevalence is determined by incidence and survival time.

- Incidence density best accounts for time at risk for disease.

- Relative risk (RR) is the ratio of incidence in exposed over incidence in unexposed. Odds ratio (OR) is the ratio of exposure odds in cases over exposure odds in controls.

- OR approximates RR when outcome is rare.

- NNT can be a clinically useful number.
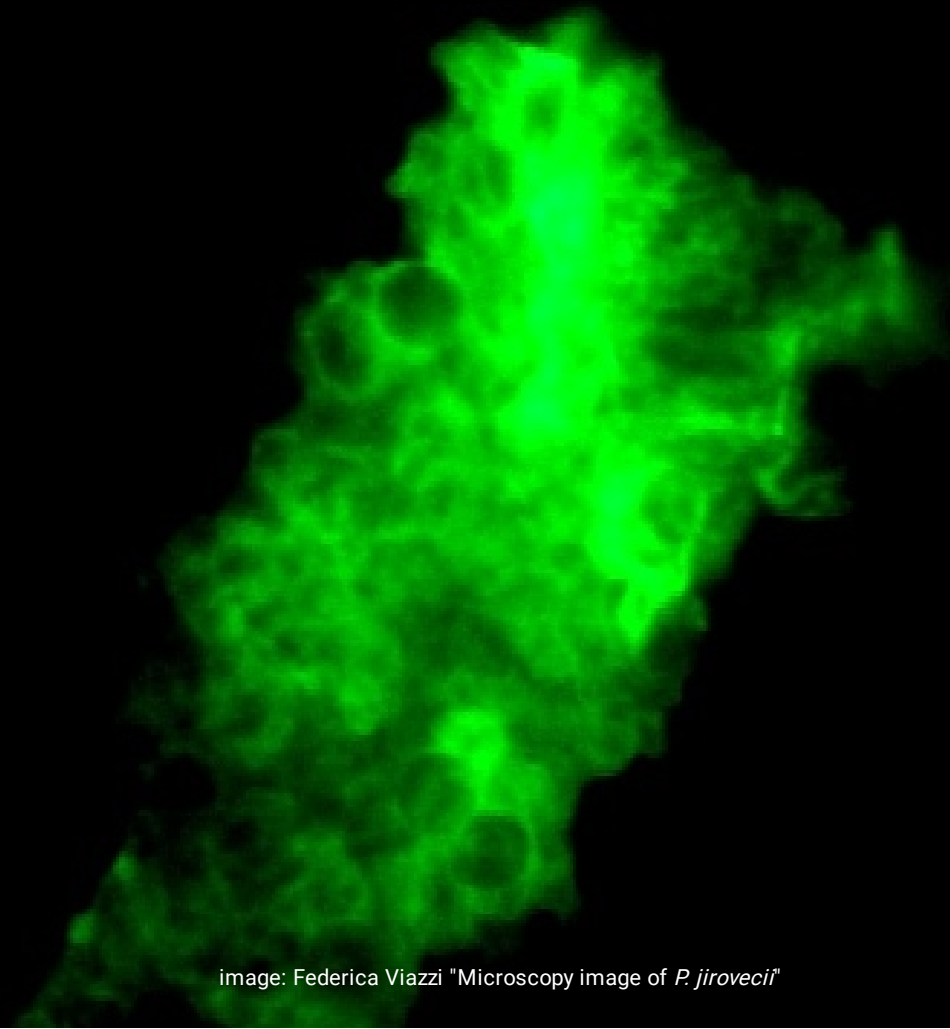
Wickramasekaran et al *Mycoses* 2017

Questions?

image: Federica Viazzi "Microscopy image of *P. jirovecii*"