Analysis in Epidemiology

Contact Information & Course Website

- Office: Allen Center 311
- Office Hours: By Appointment
- Email: Eric.Lofgren@wsu.edu
- Zoom Policies

Course material will be posted on GitHub at: https://github.com/epimodels/EpiSpring21

Generally, I will try to have material up prior to the start of class.

Purpose of this Course

- Introduce you to the basic concepts and tools of modern epidemiology
- Assist you in building a functional vocabulary for causal and systems thinking, regression models and systems models
- Let you engage with the literature, work on your own problems, etc.
- Know when, where and how to get help

Some Resources for the Curious

- *Modern Epidemiology 4th Edition*
- *Epidemiology by Design*
- stats.stackexchange.com
  - Generally speaking posting homework questions there is discouraged, but this can be a valuable resource for the future
- CISER
- The usual advice: Get help EARLY!

Center for Interdisciplinary Statistical Education and Research

A Few Philosophical and Practical Disclaimers

- There are many ways to teach epidemiology, and as a relatively new field, these continue to evolve
- Causal inference and systems thinking are often considered relatively advanced topics, to be taught well *after* an introduction to epidemiology
  - I disagree, and think we do our students a disservice by not teaching both causal and systems reasoning early
- That being said, I *work* in systems modeling, and I do not work in causal inference. This inevitably colors the material

Who Am I?

- Assistant Professor at the Paul G. Allen School for Global Animal Health
- Got my PhD from UNC Chapel Hill in Epidemiology
- Primarily interested in healthcare-associated infections, antimicrobial resistance, and emerging infectious diseases
- Unrepentant nerd, and owner of three dogs, two of which know how to open doors, one of which is at war with both the local construction crews and the local geese
- I also hate my microphone

Who Are You?

Syllabus!

Questions?

What is Epidemiology?

- Epidemiology is the study of the distribution and determinants of health-related states and events in specified populations. It is also the application of this study to the control of health problems. – – Dictionary of Epidemiology, 4th Edition
- Epidemiology is the science that studies disease occurrence and health states in human populations – Modern Epidemiology 4th Ed.
- "… epidemiology is essentially an inductive science, concerned not merely with describing the distribution of disease, but equally or more with fitting it into a consistent philosophy." - Wade Hampton Frost
- "…with respect to person, place and time."
- "Protecting Health, Saving Lives – *Millions at a Time*" – Johns Hopkins Bloomberg School of Public Health
- "The basic science of public health"

Distinctions

- Just humans?
  - Veterinary epidemiology would probably disagree
- Just diseases?
  - What is a disease?
- What is the role of epidemiology in taking action?
  - Steve Wing vs. Charlie Poole

Specific Types of Epidemiology

- "Descriptive Epidemiology" – The estimation of relationships between variables and a disease outcome
  - Often used somewhat pejoratively, and as if this was distinct from other fields
- Molecular Epidemiology – The use of (primarily) genetic tools to understand the spread and distribution of disease
- Clinical Epidemiology – Epidemiology that is specifically focused on patient populations

General Types

- “Outcome Epidemiology”
  - Cardiovascular, Cancer, Reproductive, etc.
- “Exposure Epidemiology”
  - Environmental, Social, Occupational
- Just Weird
  - Infectious Diseases

Dependent vs. Independent Happenings

- Ronald Ross in 1916
- Independent Happenings: “The frequency of the happening is independent of the number of individuals already affected”
- Dependent Happenings: The frequency of the happening is dependent on the number of individuals already affected
  - Outcome is exposure
- This is typically the Chronic vs. Infectious Diseases divide, but this line is fuzzier than it first appears
- Network effects in chronic disease outcomes

Intersections

- "…to the control of health problems"
  - Health policy, operations research
- Economics
  - A source of both exposures and outcomes, understanding tradeoffs and decisions, and a lot of shared methods (but not shared vocab)
- Ecology
  - Especially around infectious diseases
- Behavioral Science
- Statistics
- Medicine

What Makes Epidemiology Unique

- Inherently interdisciplinary
  - People arrive at epidemiology from all sorts of different paths
- Combination of methods and subject matter expertise, often in different ratios
- Epidemiology as a bridging science

How is Epidemiology Different from Biostatistics?

- All of this is opinion based
- "I can make a salad if you give me the ingredients, but I don't know how to make a good salad." – Biostatistician
- Biostats covers a wider range of disciplines, but is primarily focused on the development and refinement of statistical tools
- Epi *uses* biostatistics to uncover exposure-disease associations, but is rarely concerned with the mechanics of how that happened
- This distinction can get blurry to the point of uselessness

Questions?

Measures of Disease

- There are two major ways to measure disease
  - Prevalence
  - Incidence
- From now on, when I say disease, I also mean non-disease outcomes of interest by and large
- Prevalence is how much disease is present in *a specific population* in *a specific span of time*
- Incidence is how many new occurrences of disease took place in *a specific population* in *a specified span of time*

All Studies Are Cohort Studies

- You just might not know the cohort, or be able to sample it all
- A cohort is a group of people defined *prior to the onset of disease* who then experience outcomes over time
- This can be a generic group
  - All infants born in Denmark in 2018
- This can be extremely specific
  - All U.S. Army M1A1 Abrams tank crewmen who served as loaders during Operation Desert Storm

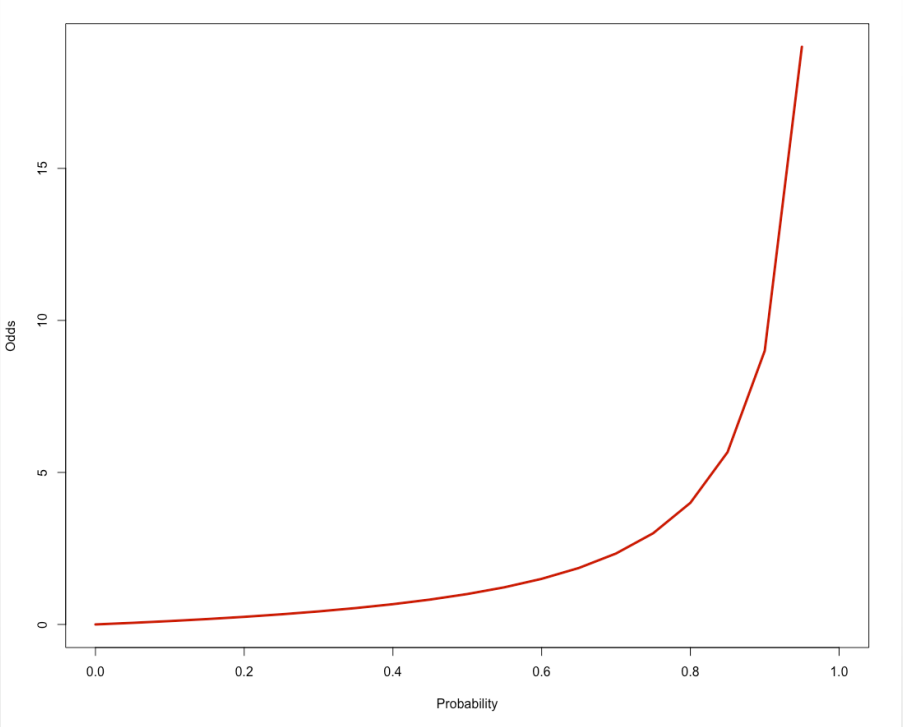
Closed and Open Cohorts

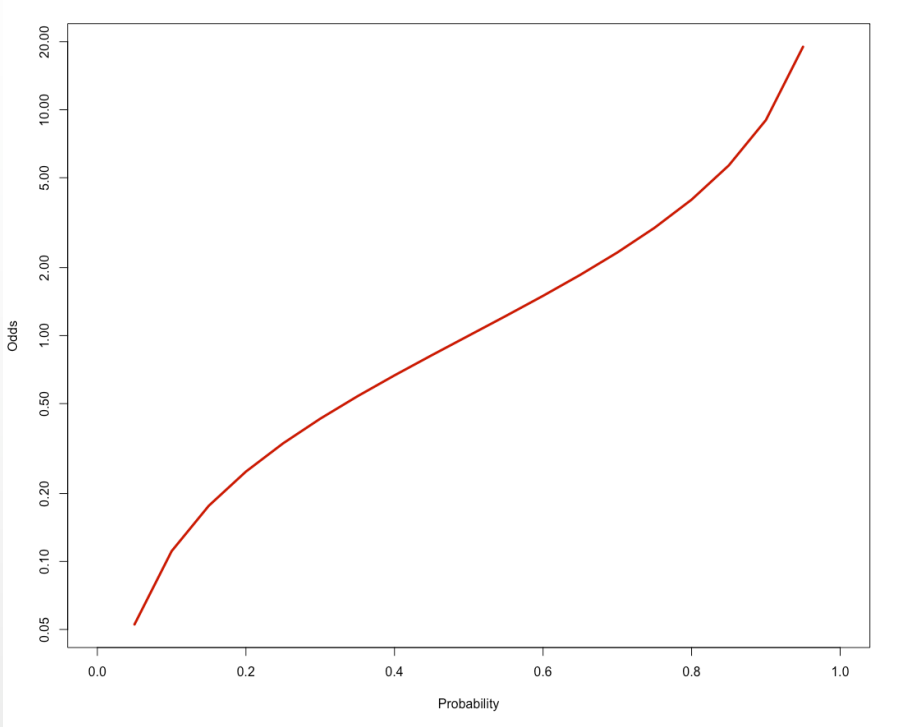
- Open Cohort: A cohort that can have members enter and leave the cohort
  - "WSU students"
  - You don't start following everyone at the same point in time
  - May get larger or smaller over time
- Closed Cohort: A cohort that can only have members leave
  - Babies born in Sweden in 1998
  - You are following everyone at *a* same point in time
  - May only get smaller

Odds

- Consider some probability P
- The odds of $P = P/(1-P)$

P	Odds
0.05	0.05/0.95 = 0.053
0.10	0.10/0.90 = 0.111
0.50	0.5/0.5 = 1 (i.e. 1:1 odds)
0.80	0.8/0.2 = 4 (i.e. 3:1 odds)


Odds
0
5
10
15
0.0
0.2
0.4
0.6
0.8
1.0
Probability


Odds
0.05
0.10
0.20
0.50
1.00
2.00
5.00
10.00
20.00
0.0
0.2
0.4
0.6
0.8
1.0
Probability

Probability vs. Odds

- When probability is low, the odds are a decent approximation
  - This stops at about $P = 0.10$ for *all strata*
- P is bounded by 0 and 1, Odds by 0 and infinity
- *Why on earth do we use odds?*
  - It's not because they're intuitive
  - They're often easier to estimate, and have some desirable statistical properties, some of which will be discussed later in this course

Prevalence

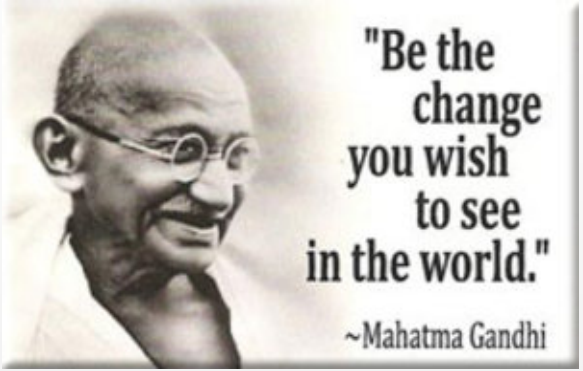
- To what extent an outcome is present in a population
- “How many people are currently living with chronic obstructive pulmonary disease in Whitman County?”

Point vs. Period Prevalence

- Point Prevalence: Prevalence at a given moment in time
  - This is never realistic – we cannot sample everyone at the same instant in time
  - Exception: Point prevalence at the beginning of a study
  - Often this is just a very small window
    - WSU's COVID-19 arrival testing
    - Sampling a dairy herd over the course of an afternoon
- Period Prevalence: Prevalence over a span of time
  - "Over the last 12 months"
  - Smaller periods cause period prevalence to converge to point prevalence
- Limits, for those who like thinking in terms of calculus

Prevalence Proportion

- The proportion of people with a given condition over a given time period
- “People with COPD in the last 10 years in Whitman County”
- Often expressed as a percentage
- Often referred to as the “prevalence rate”, despite not being a rate
  - Don’t do this



Prevalence Odds

- Prevalence Proportion/(1 – Prevalence Proportion)
- Rarely reported all on its own, though sometimes is somewhat implicity
  - “One in three women and one in four men have experienced some form of violence by an intimate partner”

Prevalence Count

- Exactly what it says on the tin – just the raw count
- This is often used when communicating with the lay public, and when the size of the population in question is well understood
- That there have been 419,220 deaths from COVID-19 in the U.S. is a prevalence count of death from COVID-19 in the U.S. population over the course of the epidemic

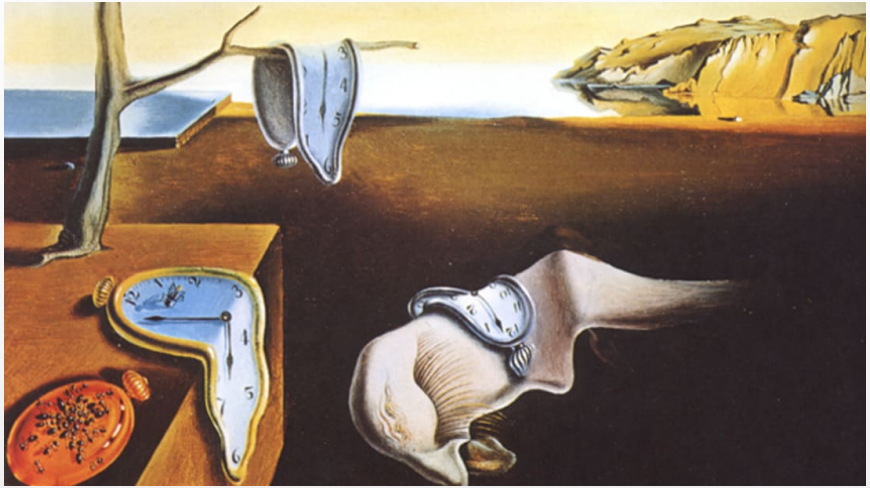
Where Prevalence Gets Tricky

- Defining your population
- Who are you talking about?
- Are you being specific?

14 Day Positive Case Count

Where Prevalence Gets Tricky

- Defining time
- There are lots of different time scales
- Point prevalence at the start of a study may mean very different ages, calendar time, etc. for everyone in the study
  - Closed cohorts follow people starting at *a* same point in time, not necessarily *all* same points in time
    - "Enrollees in the Panacea Phase III Clinical Trial" – same by study follow-up period, not the same by age, maybe not the same by calendar
    - "Babies born in Sweden in 1998" – same by age and follow-up period, the same by some calendar times (years) but not others (days, months, hours, minutes)



Pause

Questions?

Incidence

- How many new cases of a disease occurred over a given point in time
- There were 170,000 new cases of COVID-19 in the United States yesterday
- Difference between prevalence and incidence is that incidence is a measure of *occurrence* not *existence*
- Prevalence is bounded by population size – at most, everyone can have something
- Incidence is not – “Respiratory infections between 2010 and 2020”

Survival

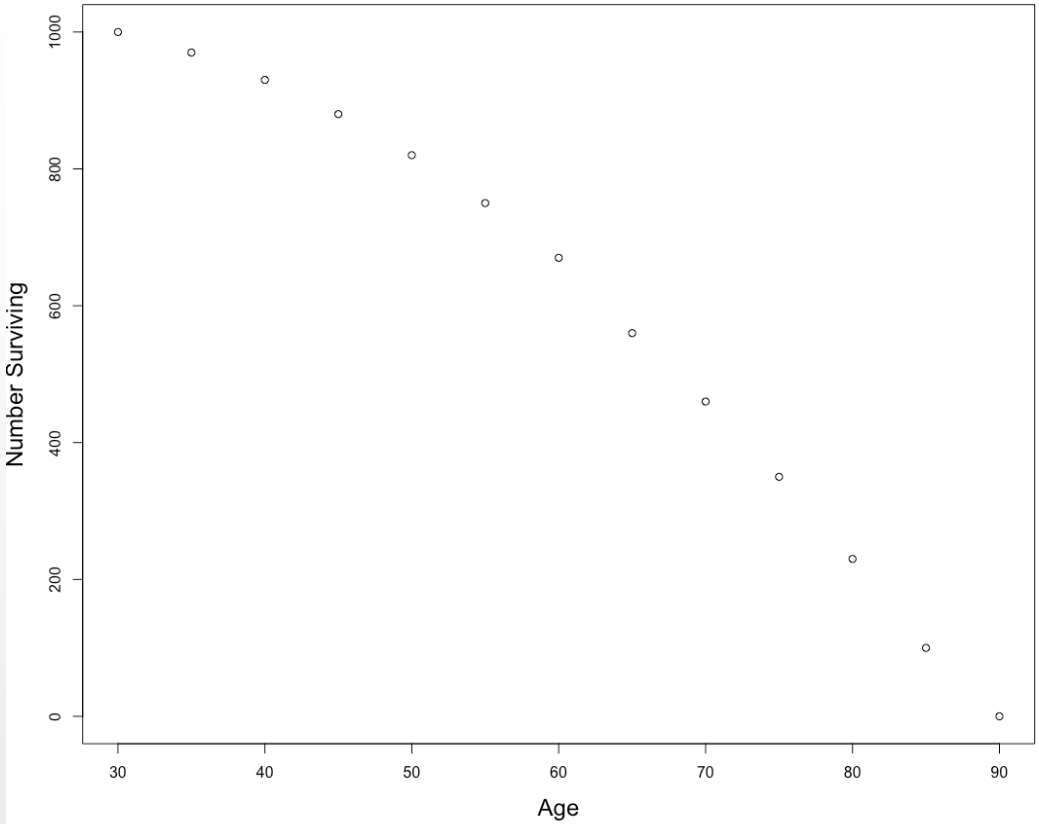
- “Survival” – Not experiencing incidence over a given period of time
- Survival is arguably the fundamental measure of most public health outcomes
  - Outcomes happen over time
  - Not relevant for sudden events – natural disasters, nuclear explosions, etc.
- We can build institution for incidence off survival

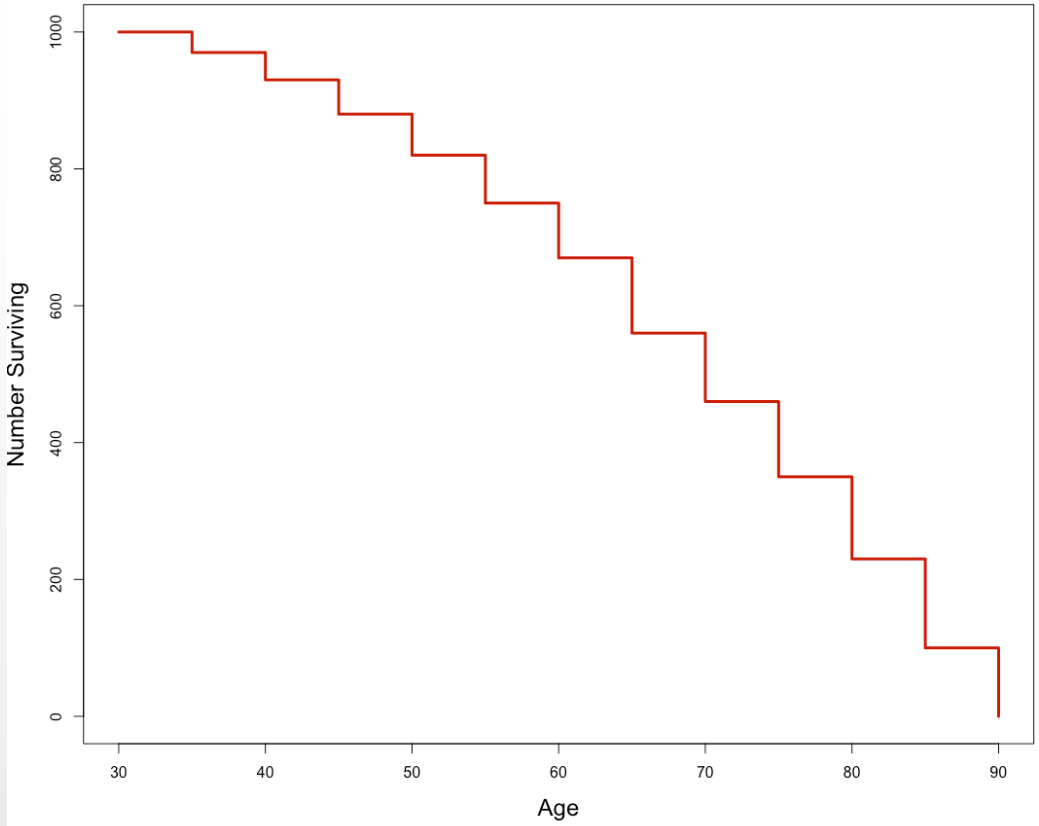
An Example

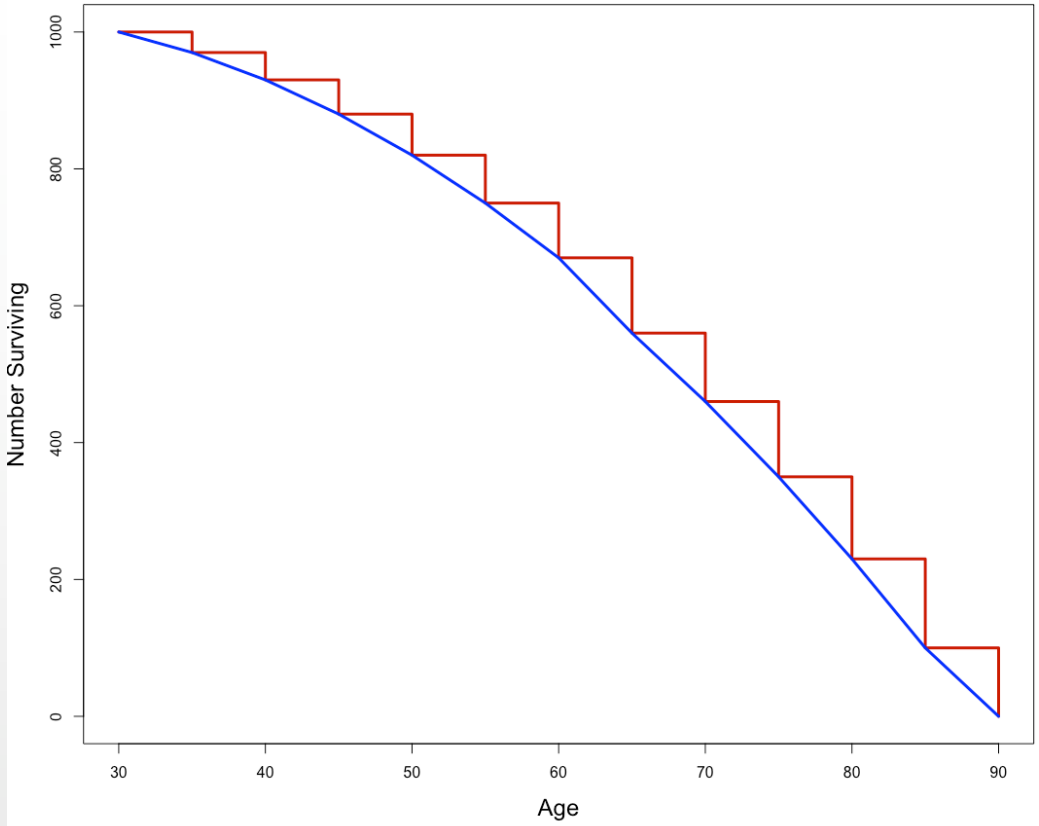
- Death is (generally) considered an unambiguous and unrepeatable outcome
- “All cause mortality” is death from any cause
- At the beginning of life, everyone (by definition) is surviving, and eventually, everyone will experience the outcome
- This makes for a very clean survival example

Survival of 1,000 Residents of Republic City

Age	Number Surviving
30	1000
35	970
40	930
45	880
50	820
55	750
60	670
65	560
70	460
75	350
80	230
85	100
90	0


Number Surviving
0
200
400
600
800
1000
30
40
50
60
70
80
90
Age


Number Surviving
0
200
400
600
800
1000
30
40
50
60
70
80
90
Age


Number Surviving
0
200
400
600
800
1000
30
40
50
60
70
80
90
Age

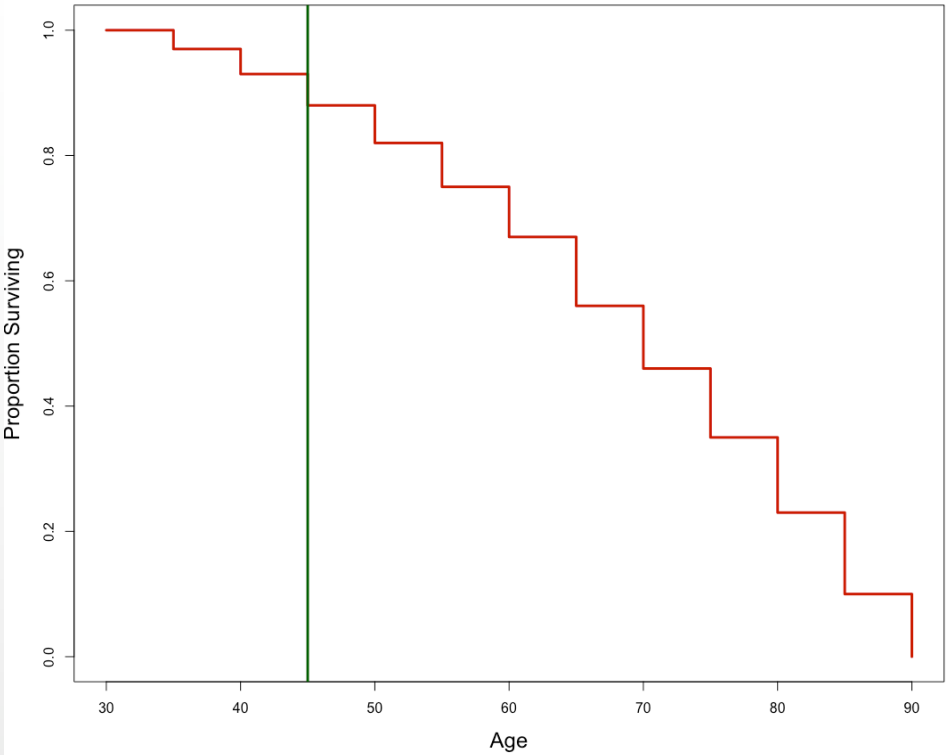
Why A Step Function?

- Mortality is probably a continuous function*
  - There are some days with higher mortality than others
  - Post-Chinese New Year is an oft cited example
- We don't have information (based on our table) at a higher resolution than one year – we basically have to assume everyone who lived past 45 but not past 50 all died at the same time
  - There are essentially infinite other scenarios that all fit the data equally well
- This is known as "interval censoring" – we'll discuss this more in a later class

Incidence Proportion

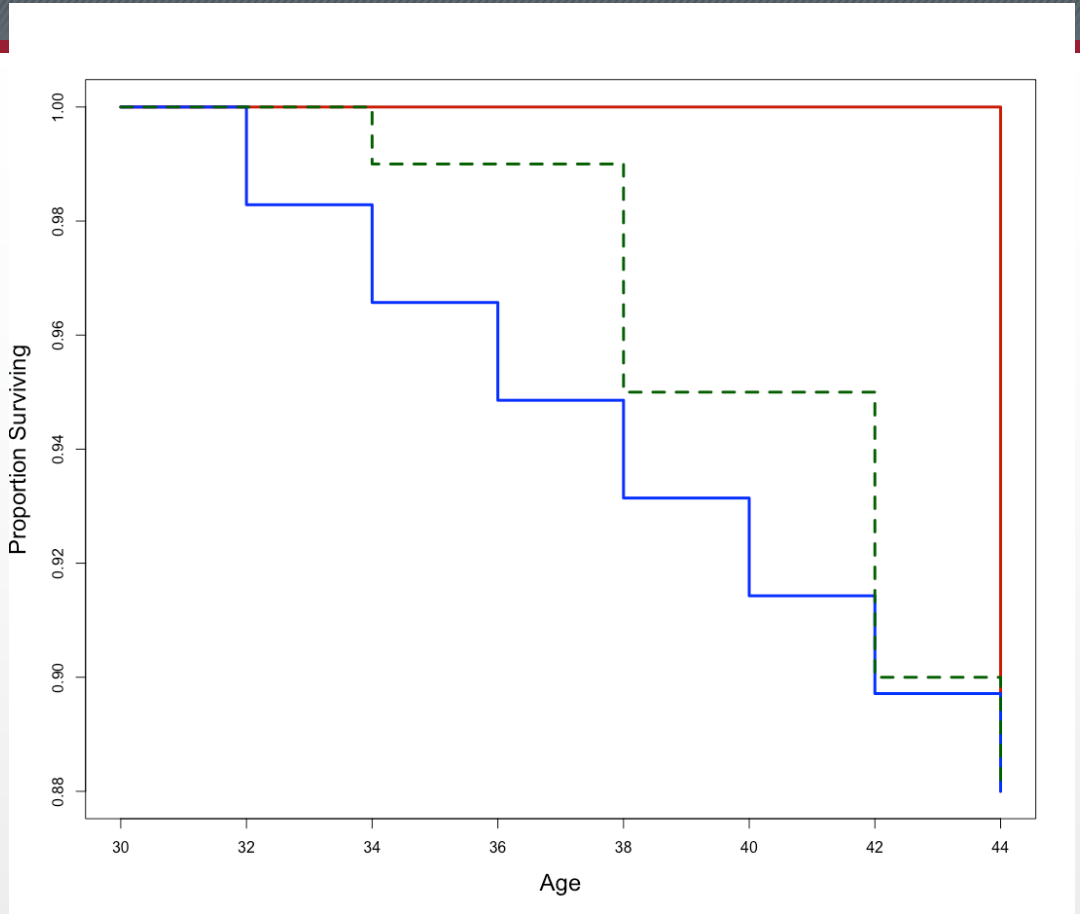
- The proportion of individuals in a population who experienced an outcome over a given span of time

- Incidence proportion of death among 1,000 residents of Republic City at Age 45
- 880 surviving to 45 years of age
- 120 have died
- 120/1000 = 0.12
- When epidemiologists use the word “risk”, this is what we’re referring to



But I Didn’t Need a Survival Curve for That

- You could indeed have just known that 880 people survived and calculated from there without ever seeing the survival curve
- This is how…most of epidemiology teaches this
- But consider…



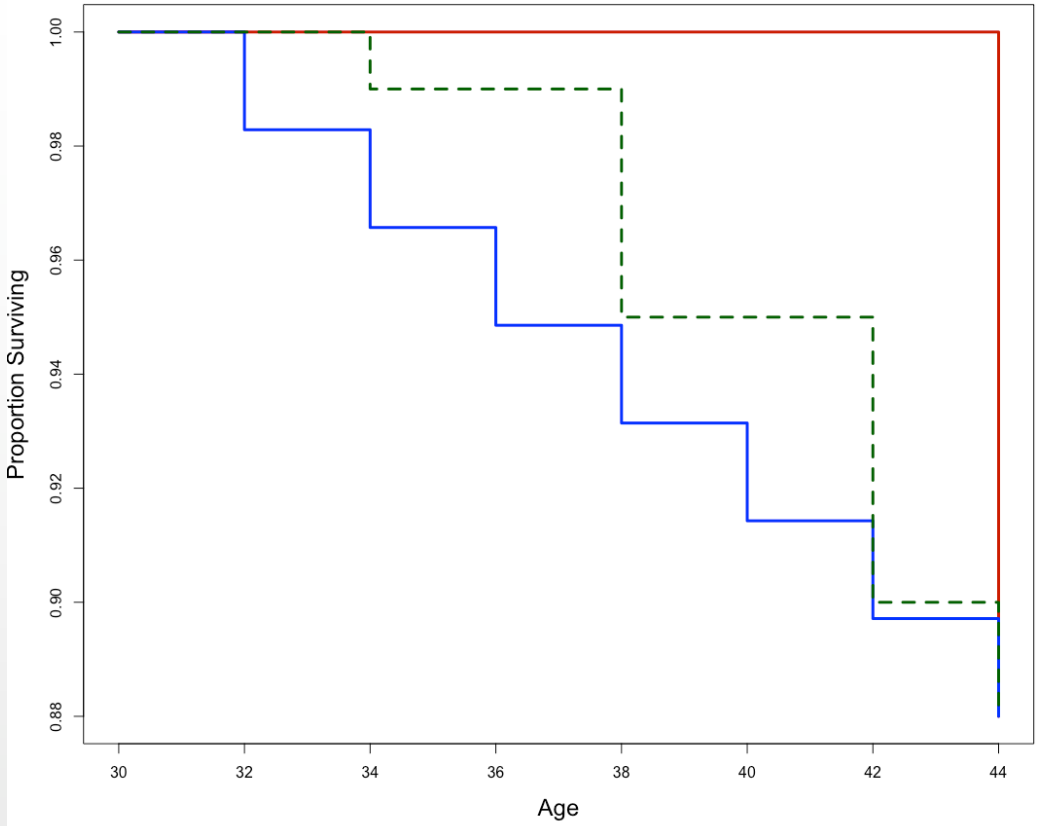
A crude incidence proportion is a low-fidelity look at the underlying real survival curve

Incidence Odds

- Incidence Proportion/(1-Incidence Proportion)

Incidence Rate

- Genuinely a true-rate
- Events/Person-Time
- Person-Time = People Followed x Time Followed
- Four subjects followed for 6 months = 4x6 = 24 person-months
- Note this scales – in theory, one person followed for 24 months should have an equal number of events as four people followed for 6 months, or 24 people followed for one month

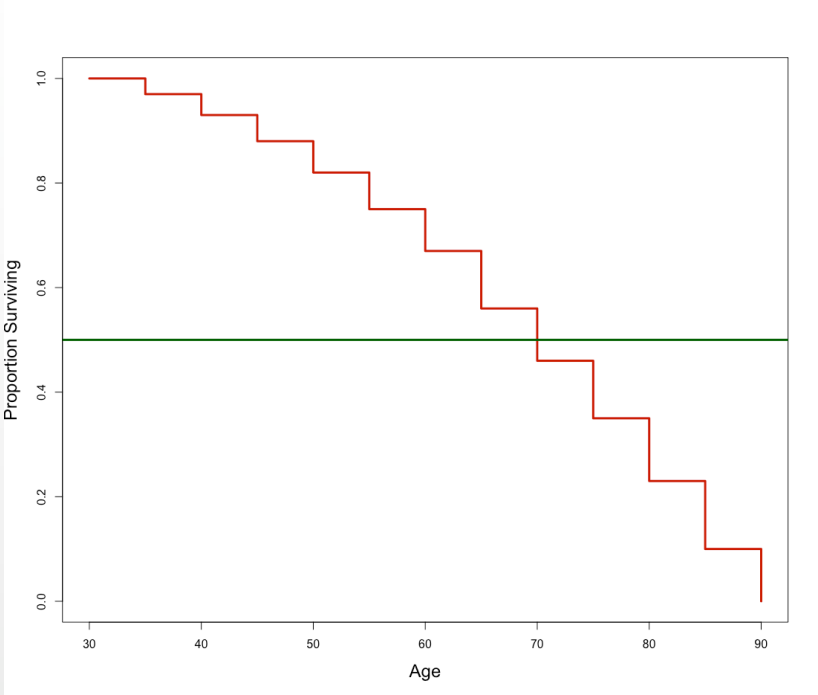

Proportion Surviving
1.00
0.98
0.96
0.94
0.92
0.90
0.88
30
32
34
36
38
40
42
44
Age

Other Nice Features of the Incidence Rate

- Does not need to be reported per a specific span of time
- Scale is easy to adjust. Given an incidence rate of 2 deaths per 1000 person-years, you can predict the number of deaths in a city of 276,428 over a 4.76 year period.

Time to Event

- We have estimated the risk of an event occurring at a particular time
- We can also estimate things like the median time to an event
- This is underused in epidemiology
- We will discuss how to estimate this statistically in a later class



Where Incidence Gets Tricky

- Repeating events
  - Are you only counting the first event, or all events
  - When someone experiences an event, are they immediately susceptible to a repeat, or is there some time?
    - This impacts everything from pregnancy to infections
- Study dropout
  - You can't see events occur in people who have left
  - What if those people left *because they experienced an event?*
- Competing risks
  - What if participants experience an event that is mutually exclusive with the event you're looking at?
    - "You can't die of cancer if you got hit by a train"

Questions?

Relationships Between Incidence and Prevalence

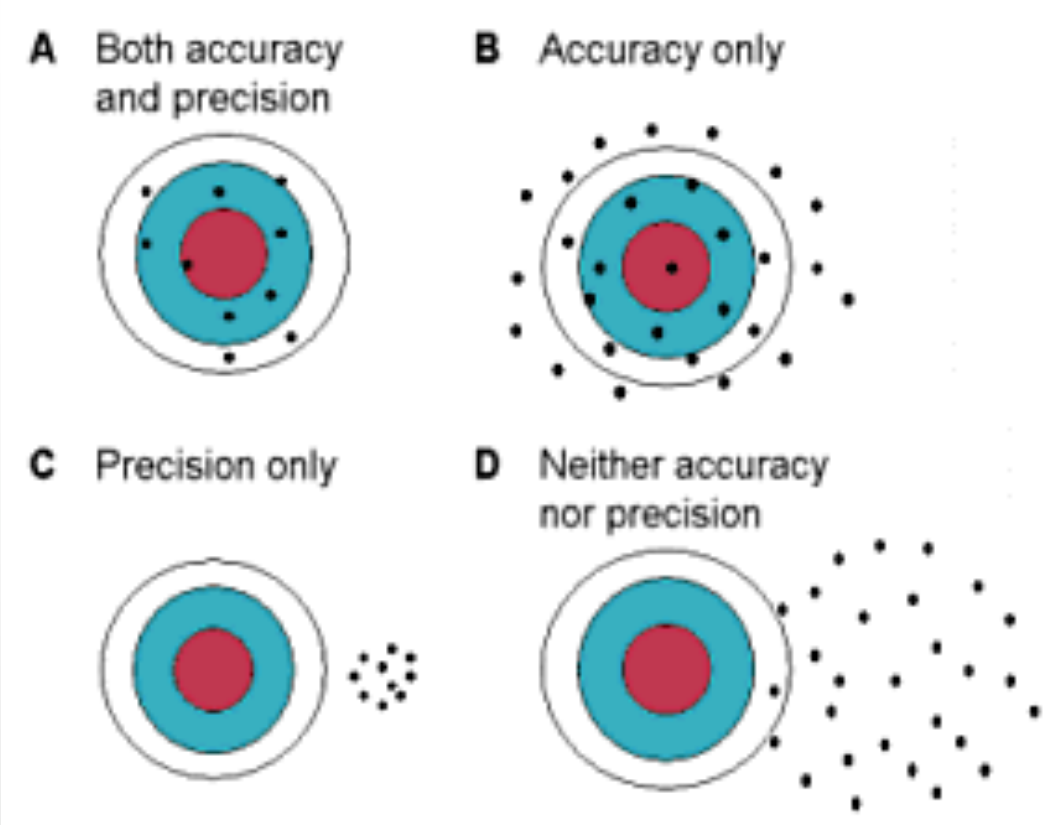
- Prevalence $\approx$ *Incidence X Duration*
- This make sense – the number of events that occur in a period and how long they last yields how many events can be measured
- The precise relationship is:
- Prevalence Odds = Incidence Rate X Average Duration of Outcome
- This can result in unintuitive results
  - Improving treatment for AIDS causes the prevalence of AIDS to rise

Risk (Incidence Proportion) and Incidence Rate

- $Risk = 1 - e^{-IncidenceRate \times Time}$
- This can be approximated with Risk = Incidence Rate x Time
- Note: This is only true in a closed cohort

Types of Error

- We will talk a lot about error in this class
- There are two (broad) types of error: Random and Systematic
- Random error is just that – there is noise in your estimate that is random
  - An estimate with high random error is imprecise
- Systematic error is when there is something wrong with how we are measuring
  - “The camera always adds 10 pounds”


A Both accuracy and precision
B Accuracy only
C Precision only
D Neither accuracy nor precision

Systematic Error and Bias

- We often use the term “bias” to refer to systematic error
- I think this is ever so slightly inaccurate for systems models, but we’ll get into that later
- For the most part, you can regard these as synonyms

Questions?