

# L1: Questions and time in epidemiologic studies

EPID 722

Spring 2021

UNC – Chapel Hill

jessedwards@unc.edu

# Roadmap

Chapter 1: 10 considerations when asking questions

Chapter 2: Risk

Chapter 3: Why learn methods for survival analysis?

Chapter 4: Line diagrams

# Chapter 1: Asking questions

```
> mydata
 [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15] [,16]
[1,] NA NA
[2,] NA NA
[3,] NA NA
[4,] NA NA
[5,] NA NA
[6,] NA NA
[7,] NA NA
[8,] NA NA
[9,] NA NA
[10,] NA NA
[11,] NA NA
[12,] NA NA
[13,] NA NA
[14,] NA NA
[15,] NA NA
[16,] NA NA
[17,] NA NA
[18,] NA NA
[19,] NA NA
[20,] NA NA
```

**1. We have 100% missing data for all questions that are not asked**

## 2. There are many routes to asking a question, and many sources of influence



How do we decide what to ask?



Who decides which questions we answer?

### 3. “Asking” a question is a 2-step process



Idea

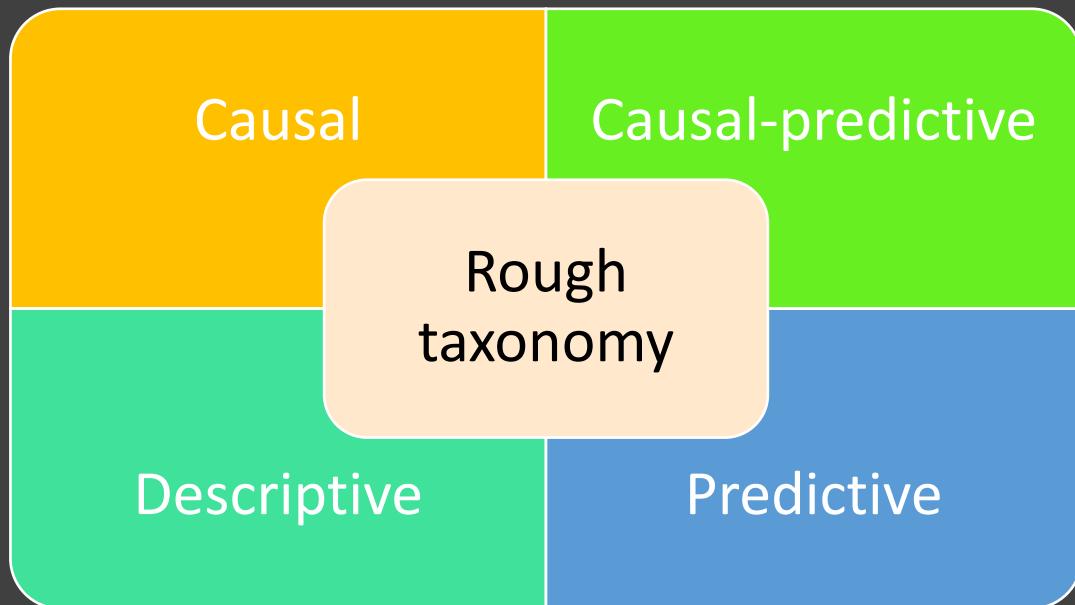
What problem  
needs to be  
addressed?



[Question]

Frame the  
question

# 4. Different types of questions can inform decisions



# 4. Different types of questions can inform decisions

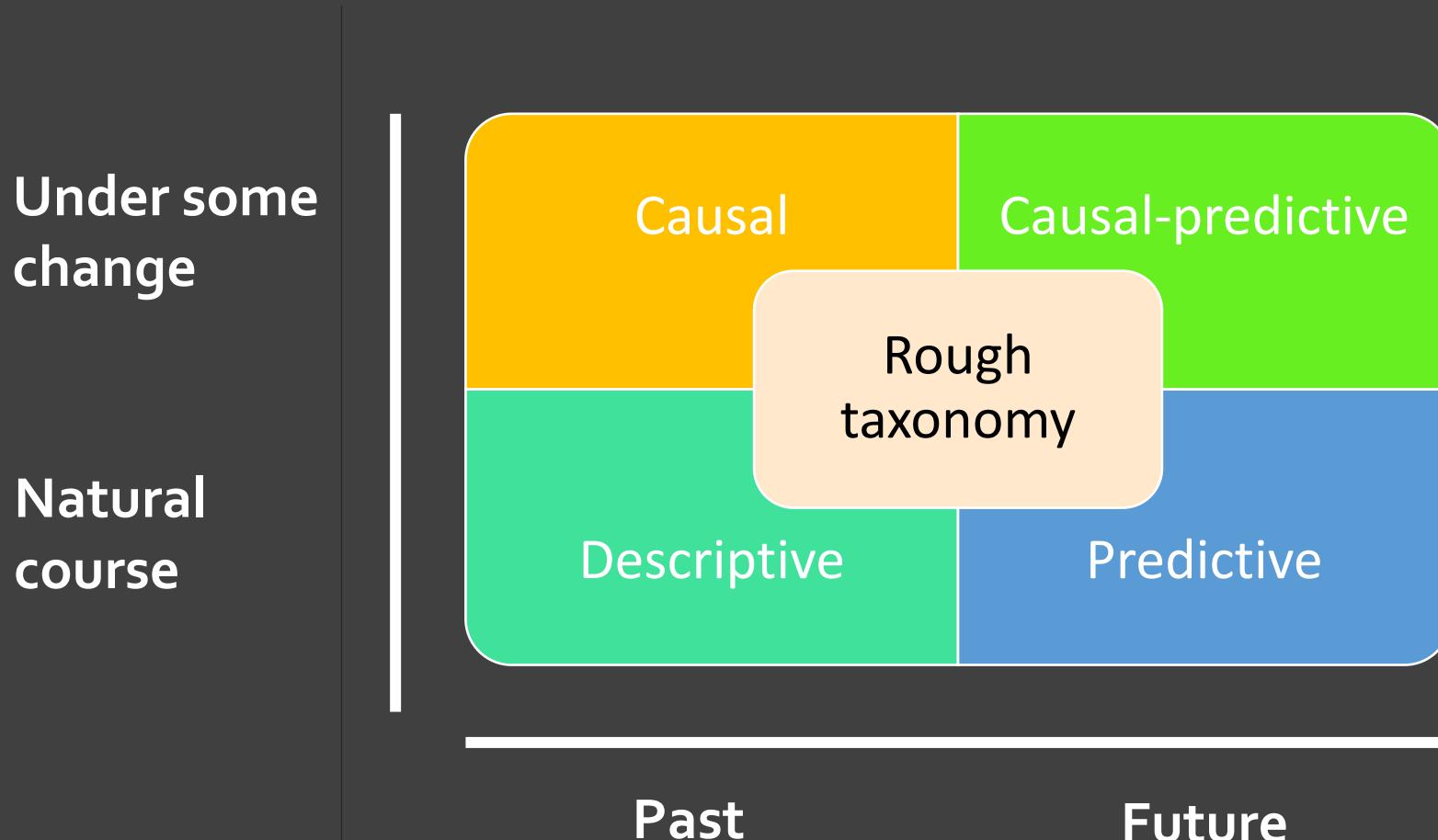
**Descriptive:** What happened?

**Predictive:** What will happen?

**Causal:** What would have happened had we done X?

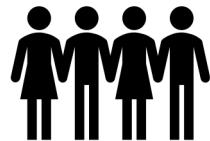
**Causal-predictive:** What will happen if we do X in the future?

# 4. Different types of questions can inform decisions



# 5. Considering the use of the results is important to framing the question

---



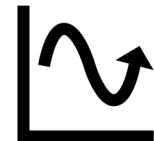
Who will be acting on the results and  
Whom does the decision affect?

---



What actions are feasible, now or in  
the future?

---



What types of results would be  
compelling to decision makers?

# 6. To maximize utility of the answers, questions should have a set of components



1. Target population



2. Time period of interest (origin and timescale)



3. Outcome(s), including WHEN outcomes should be measured



4. If there are a set of **actions** that are being assessed, the set of candidate actions, including WHEN they would occur



5. If there is a **comparison** being drawn **between groups**, the set of groups compared, including WHEN groups are defined.

# 7. Consider the idealized cohort study to flesh out your question

## A NEW APPROACH TO CAUSAL INFERENCE IN MORTALITY STUDIES WITH A SUSTAINED EXPOSURE PERIOD—APPLICATION TO CONTROL OF THE HEALTHY WORKER SURVIVOR EFFECT

JAMES ROBINS

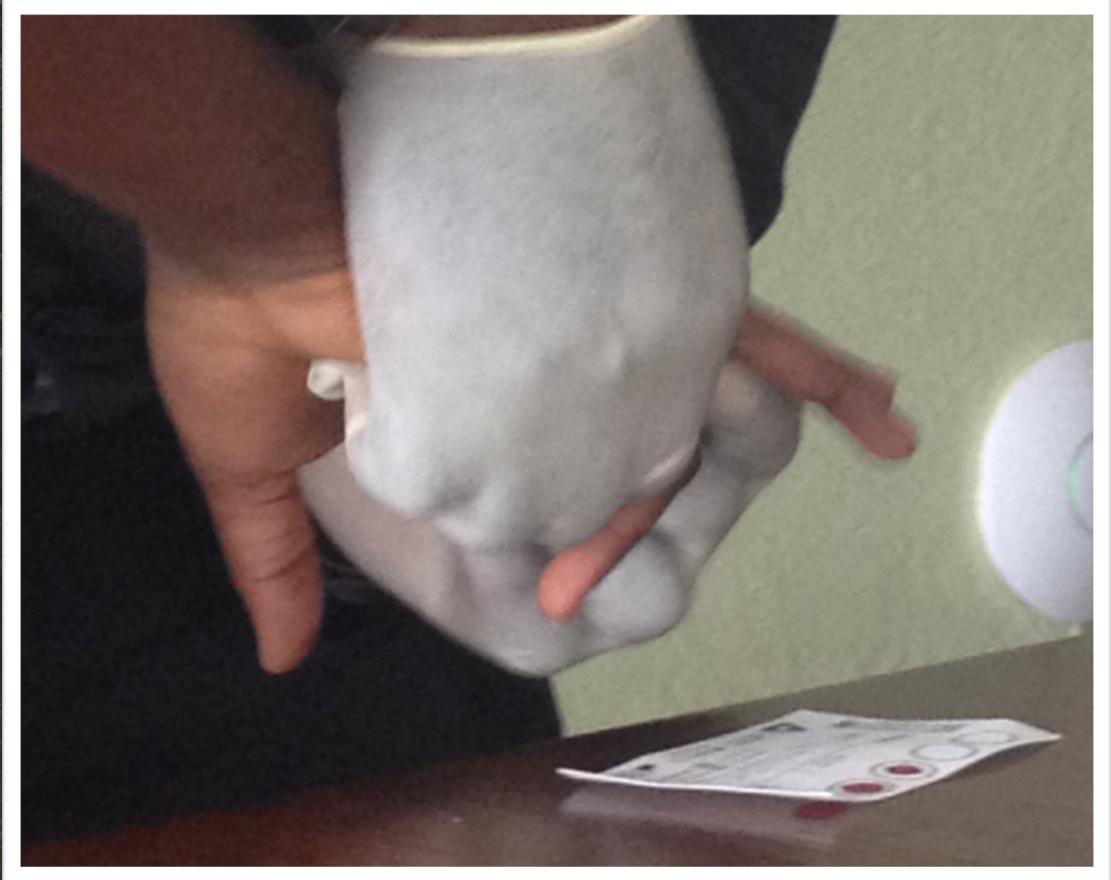
Harvard School of Public Health  
665 Huntington Avenue  
Boston, MA 02115

“Target trials” and beyond

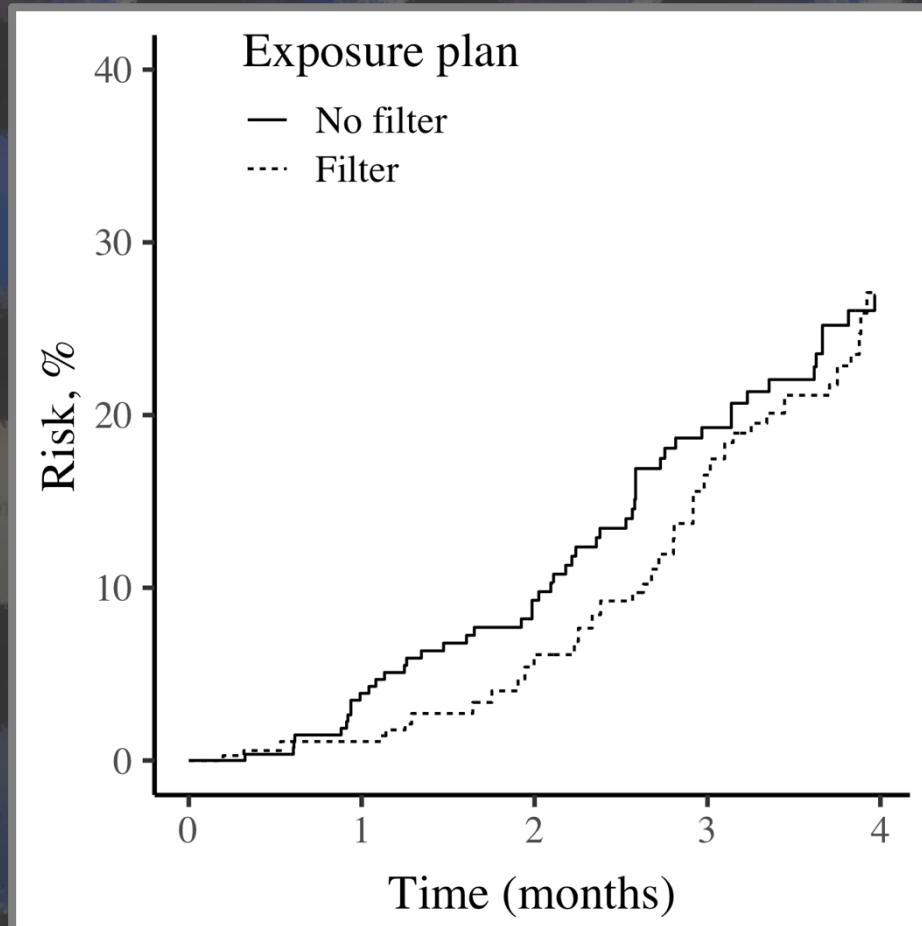
The answer to this question often not only is helpful in planning a retrospective study but also may be of assistance in determining

2. OBSERVATIONAL STUDIES AS RANDOMIZED TRIALS MISSING DATA ON TREATMENT PROTOCOL

## 8. Frame the question without considering existing data or the logistics of data collection



# 9. Consider absolute measures in addition to contrasts



# 10. Always ask “compared to what”



WHAT IS A REALISTIC  
ALTERNATIVE TO ANY  
ACTION CONSIDERED?



ARE THERE COMPETING  
EVENTS OR OTHER  
RISKS/BENEFITS TO BE  
WEIGHED



IMPORTANT FOR  
METHODS, AS WELL AS  
SUBSTANTIVE  
QUESTIONS

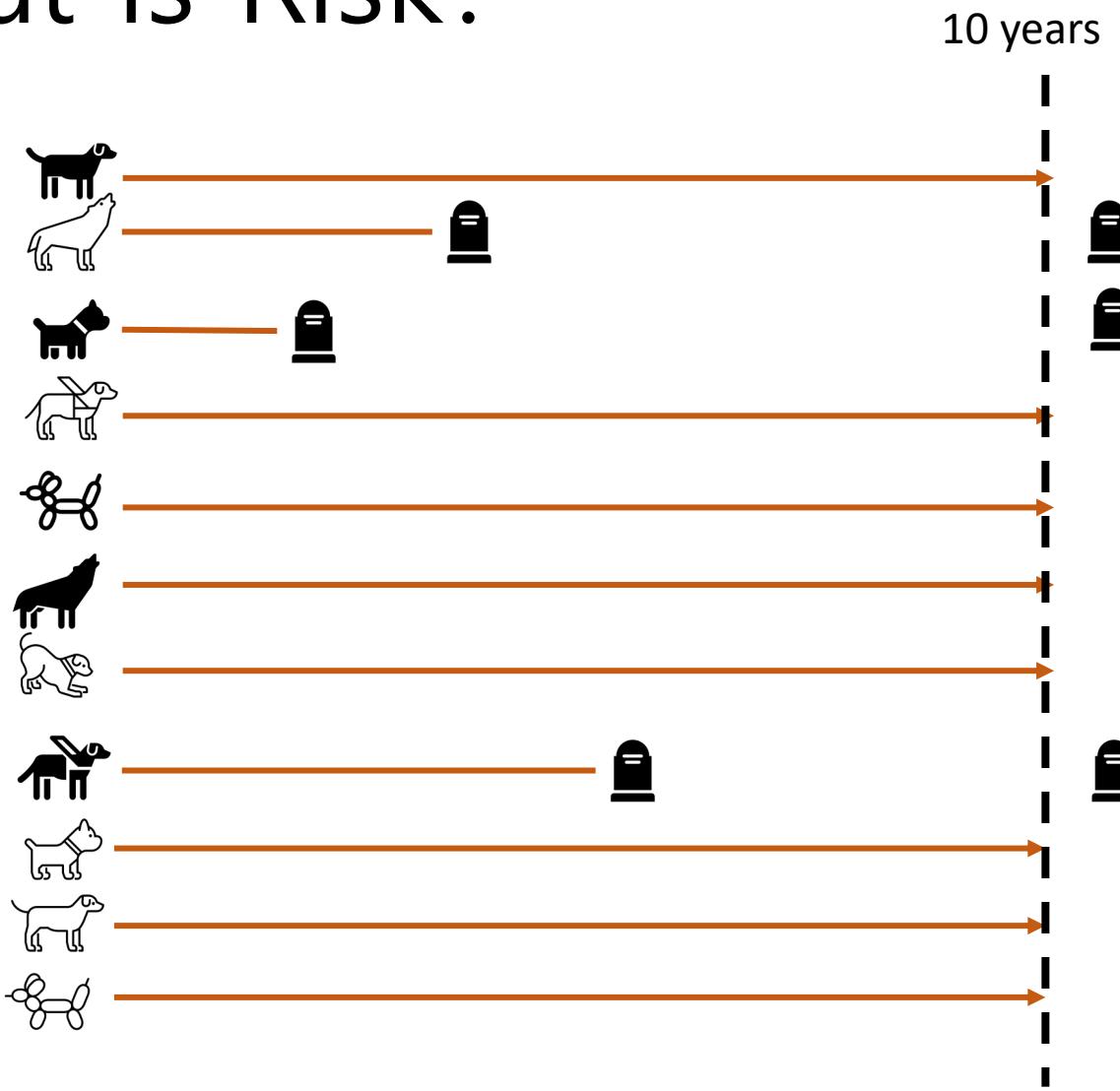
# Chapter 2: Recipe for risk in closed cohorts

# What is Risk?

In earlier courses, we defined risk as the proportion of units who experienced the outcome within a defined time period.



# What is Risk?

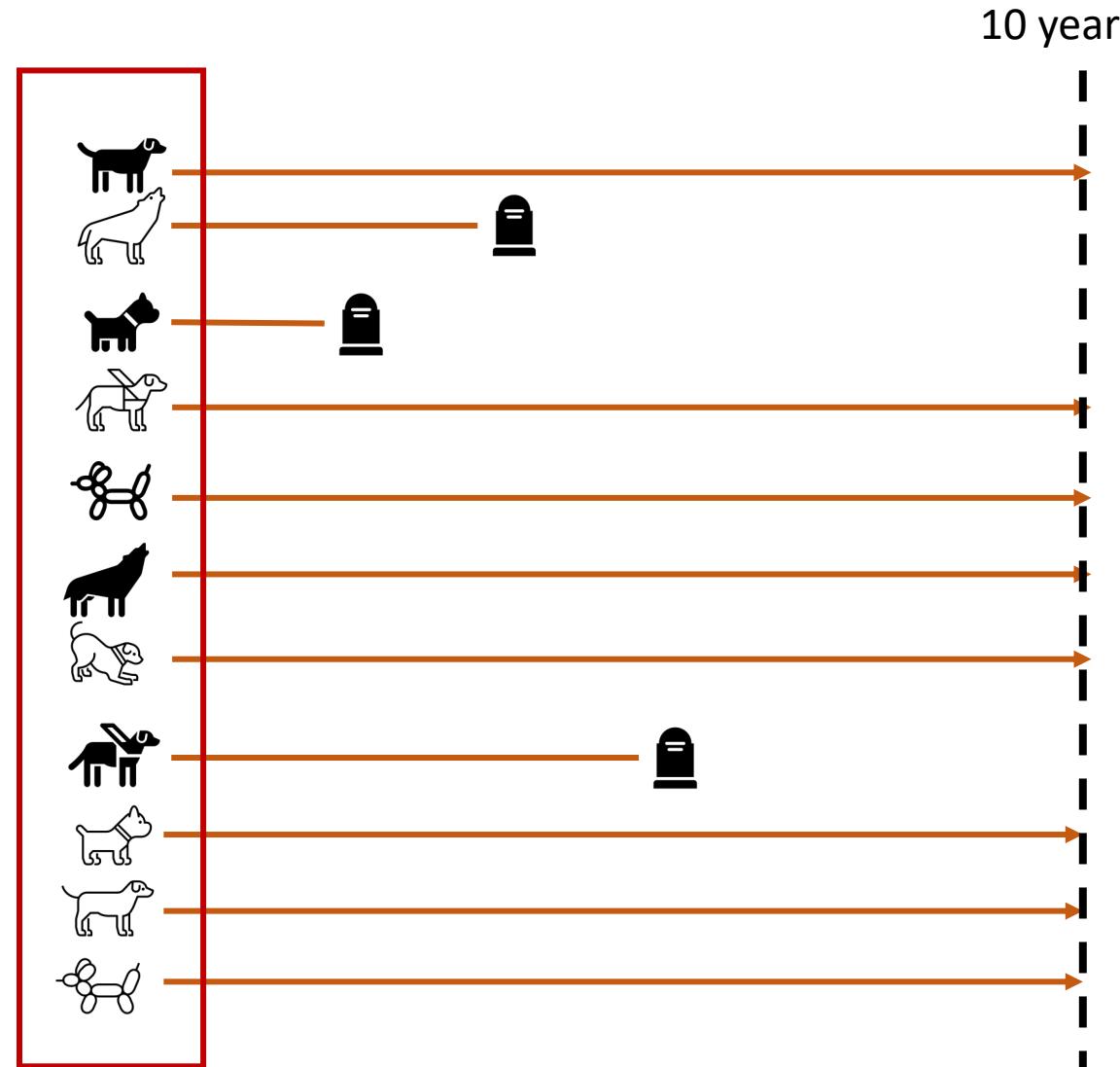


3 deaths  
11 dogs

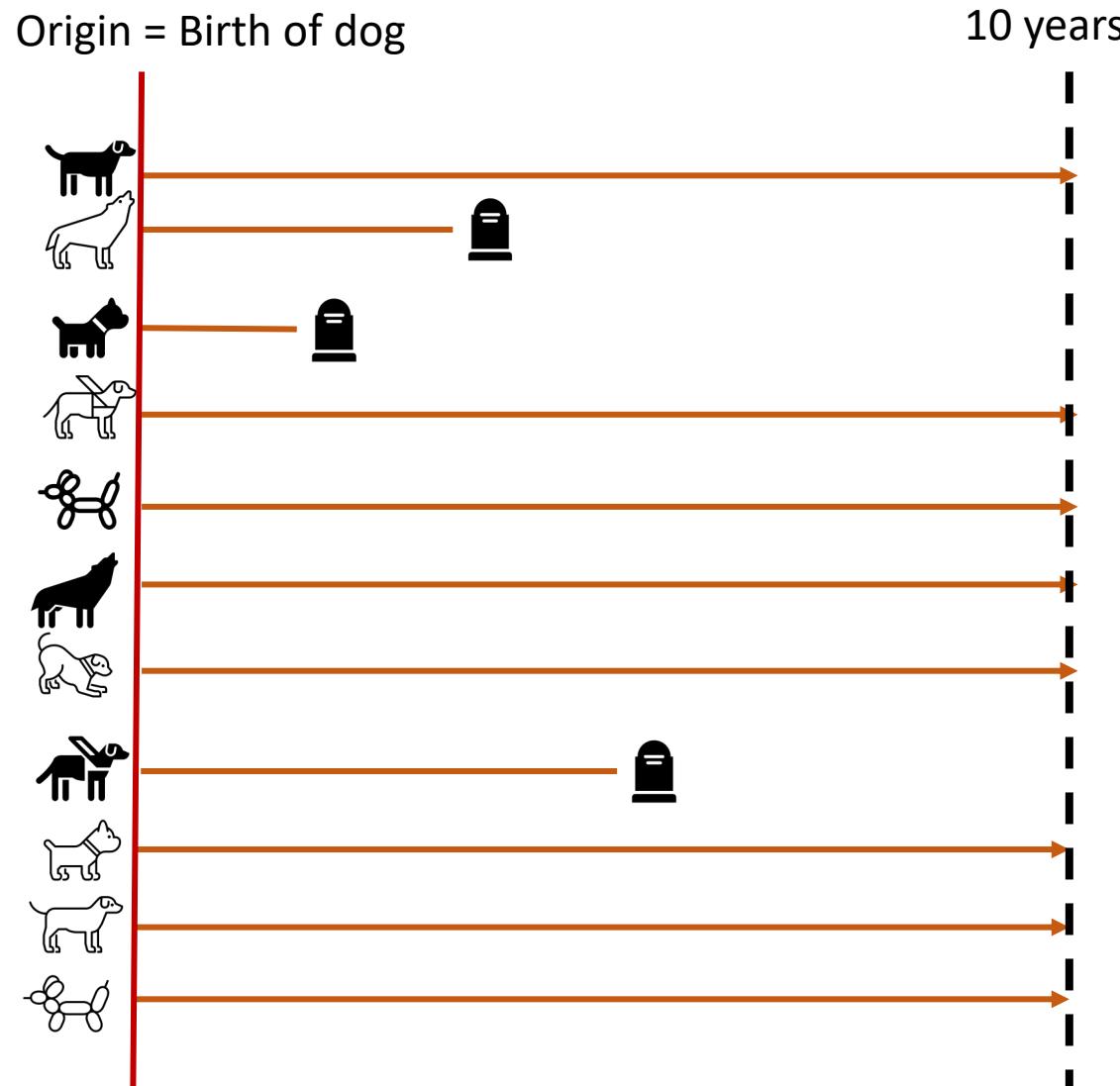
10-year risk = 27%

# Refining the question: Target population

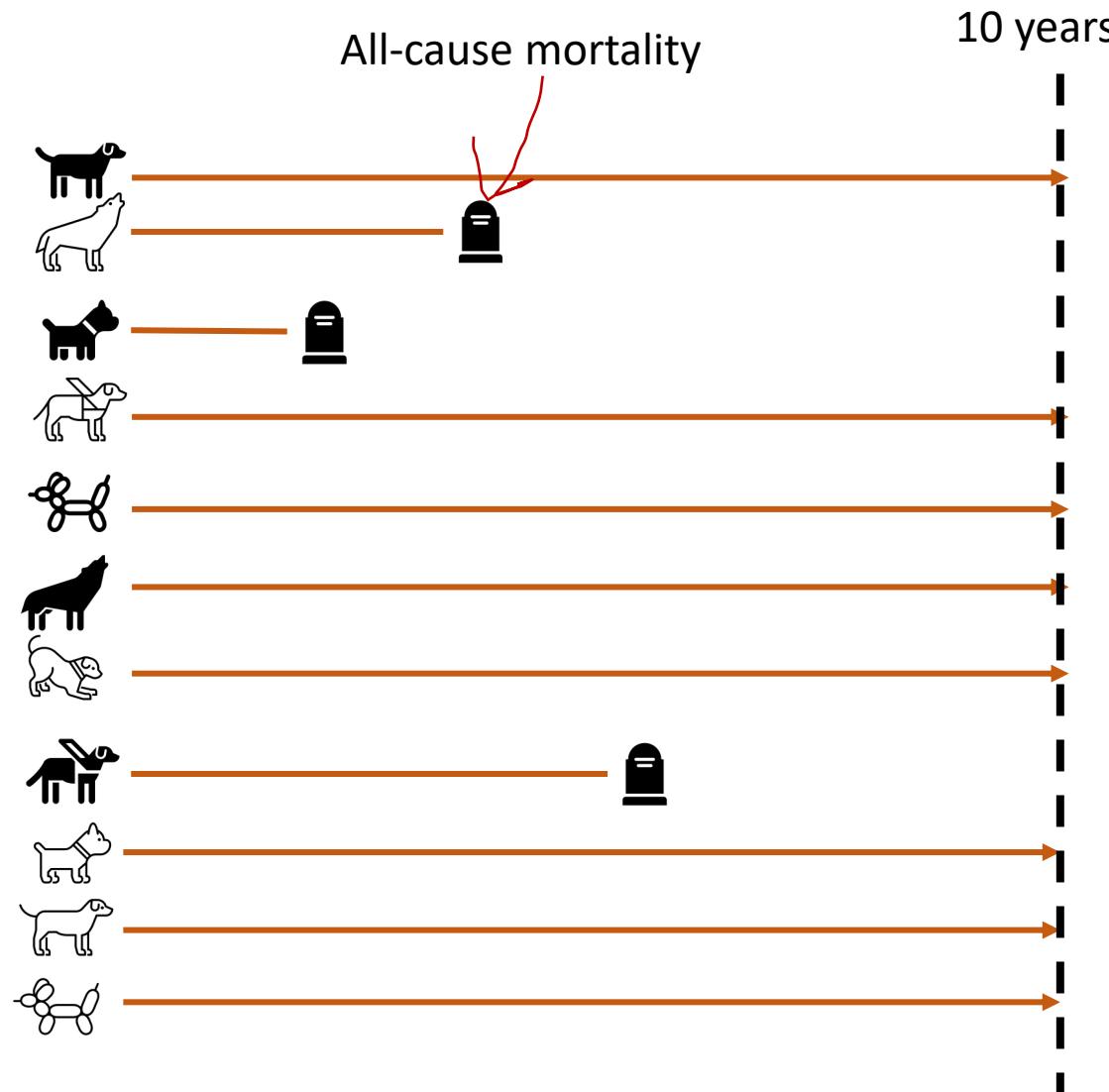
Dogs in Carrboro



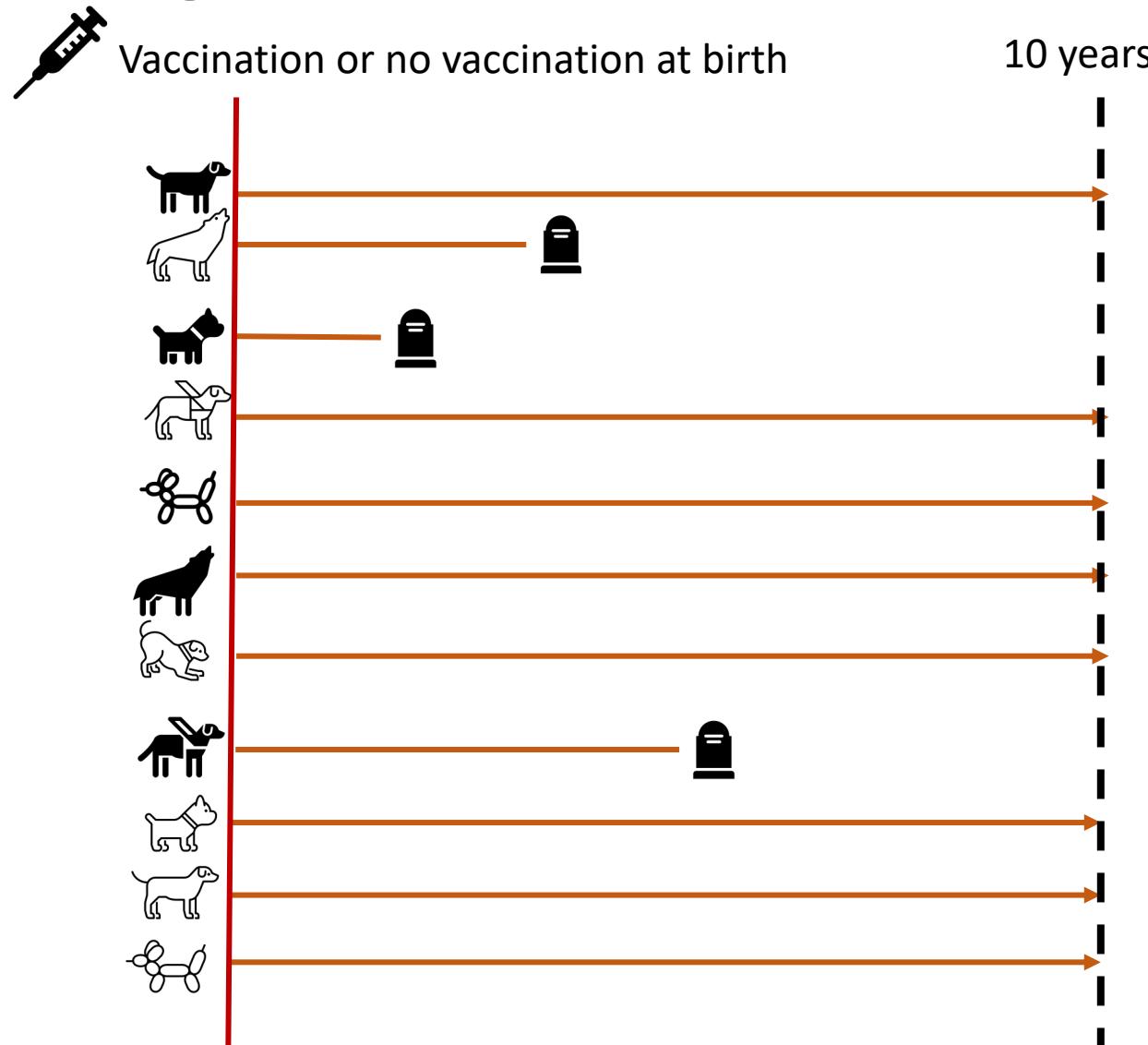
# Refining the question: Time period



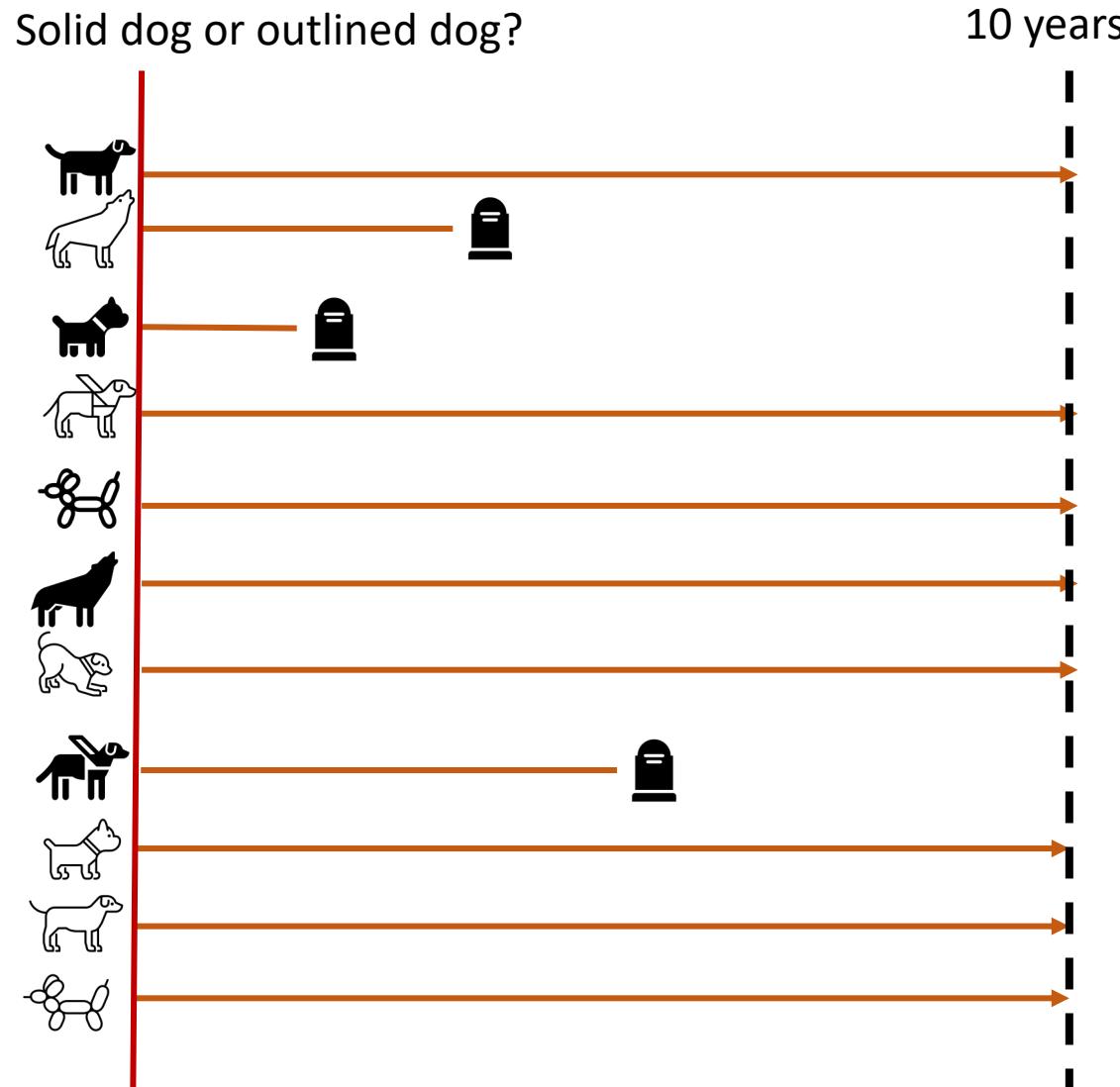
# Refining the question: Outcome(s)



# Refining the question: Actions

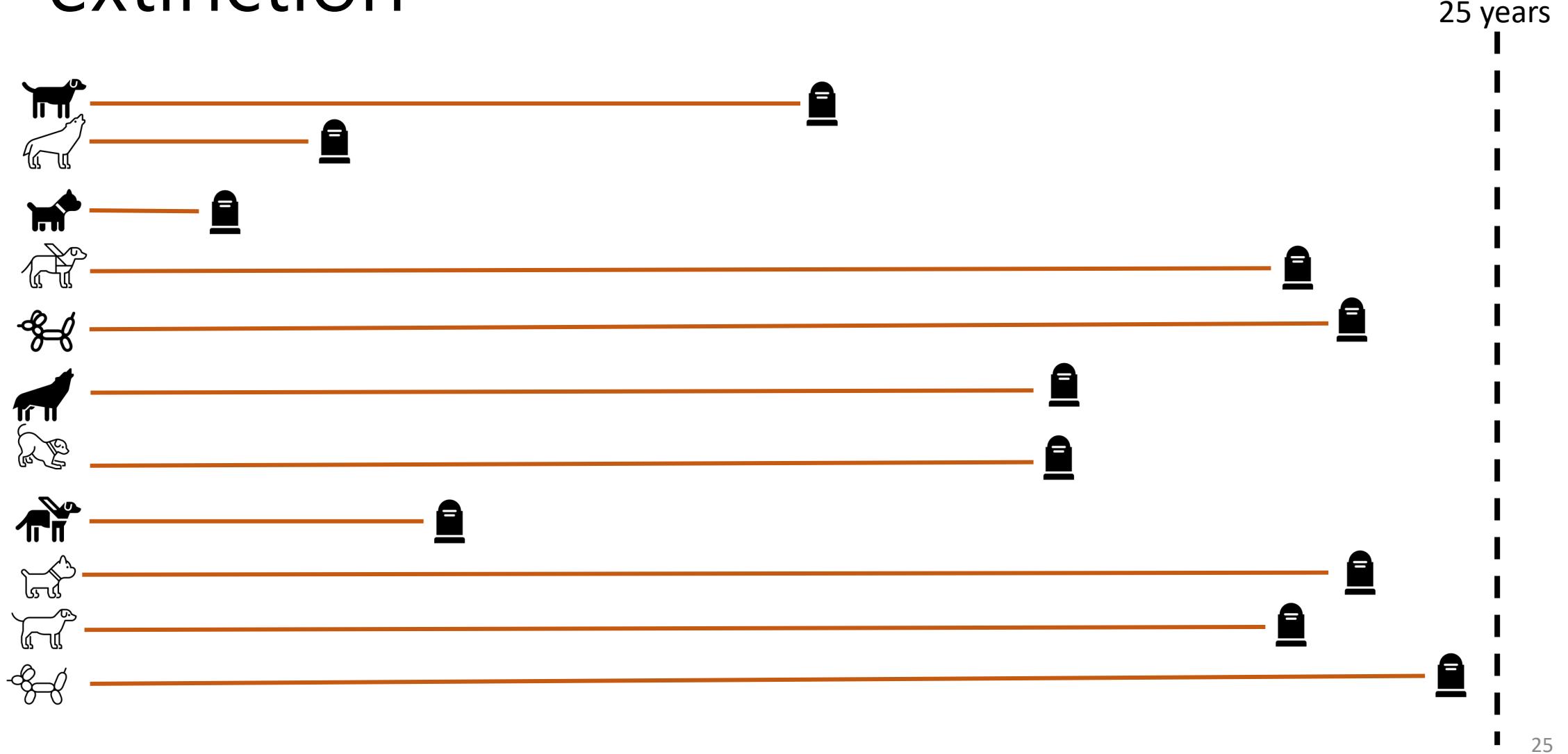


# Refining the question: Groups



# Chapter 3: Why learn survival analysis?

# Learn about cohorts followed to extinction



# Risk as a function

## First, some notation

Let  $T_i$  represent the time from the origin to death for each dog  $i$

## Risk as a scalar

Risk at 25 years =  $P(T_i \leq 25) = 1$

Or

Risk at 10 years =  $P(T_i \leq 10) = 0.27$

# Risk as a function

## **First, some notation**

Let  $T_i$  represent the time from the origin to death for each dog  $i$

## **Risk as a function**

Risk at  $t$  years =  $P(T_i \leq t) = F(t)$

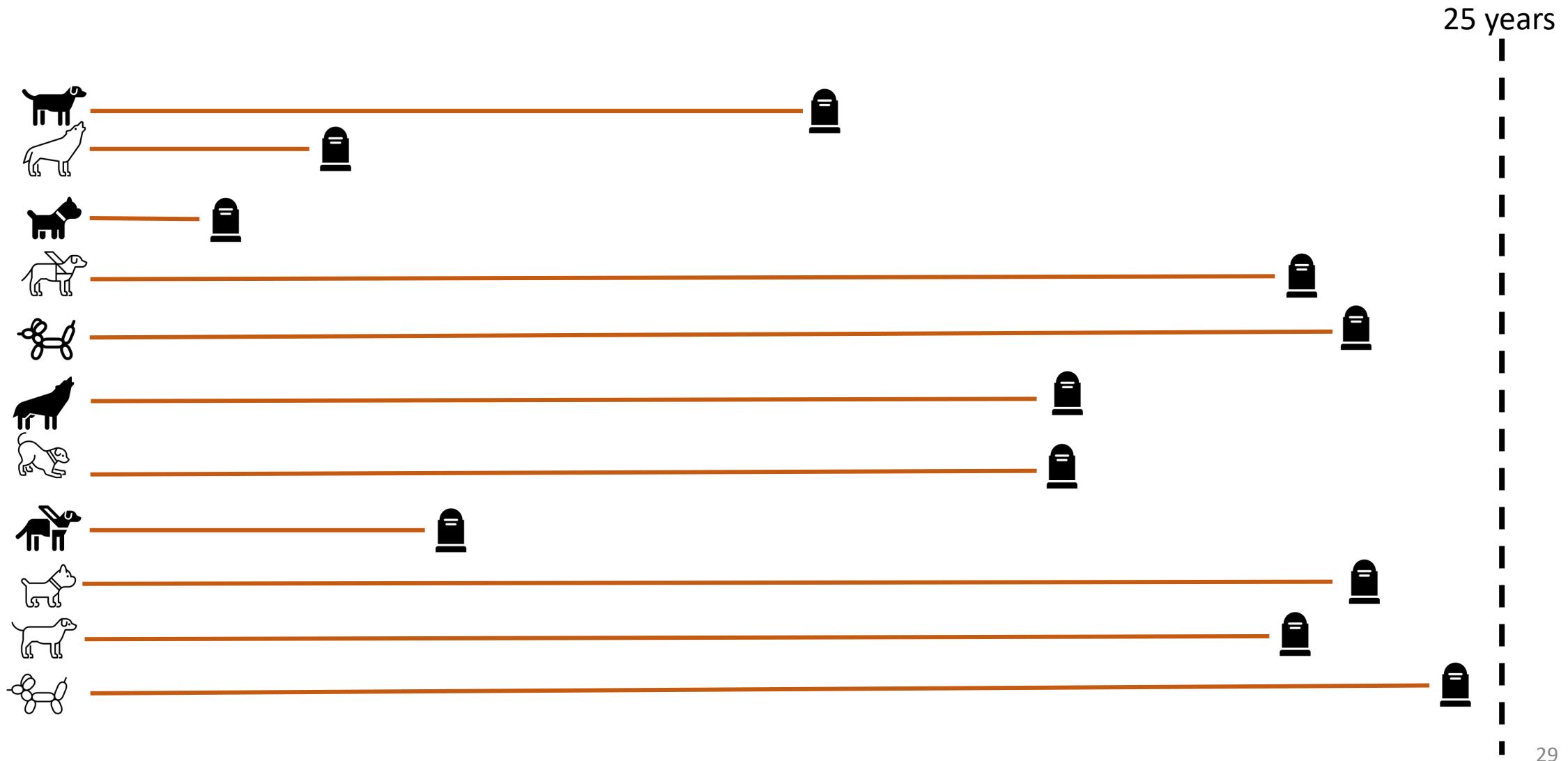
# Learn from non-closed cohorts

Cohorts may be *open on the right* or *open on the left*

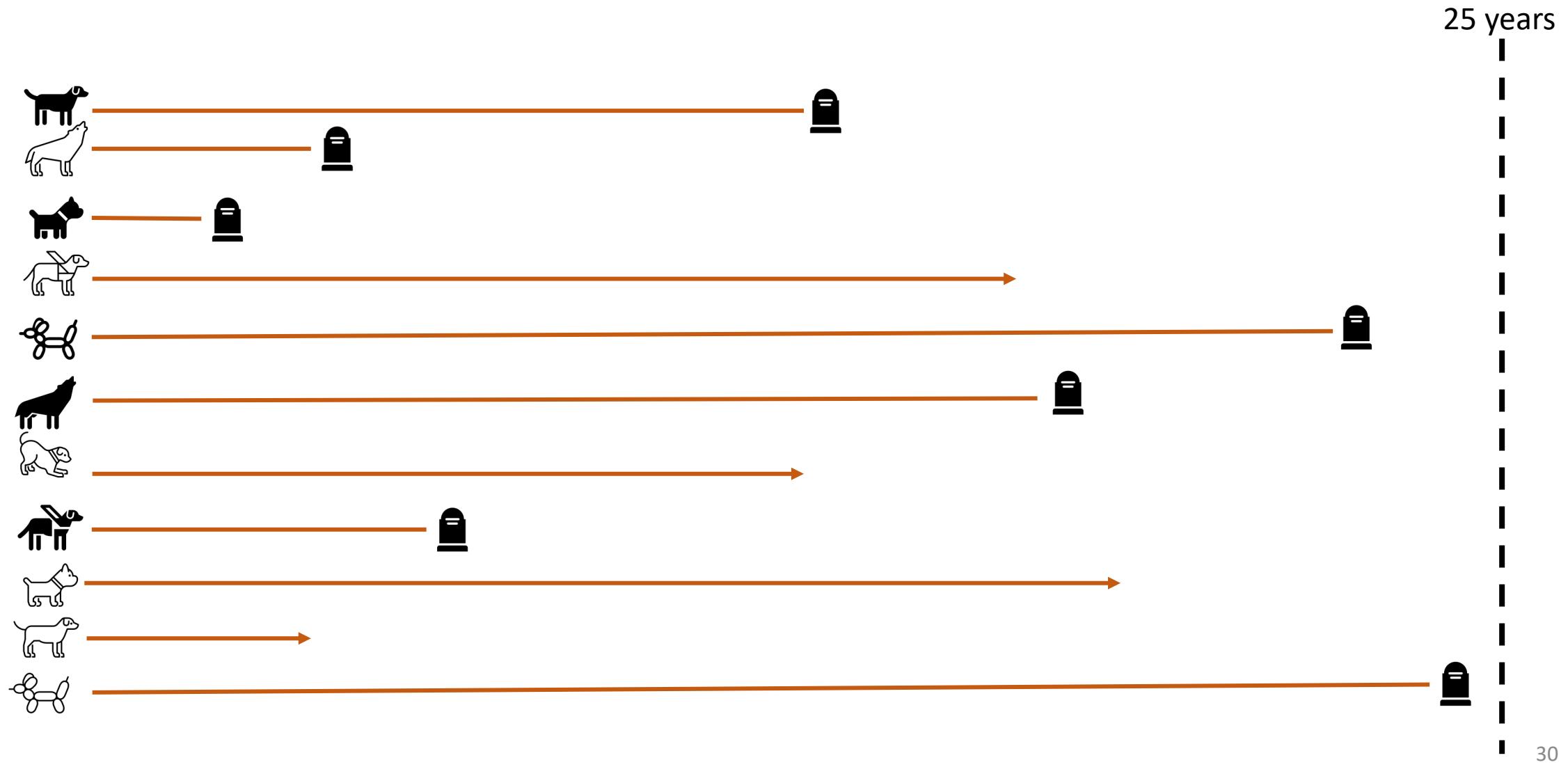
Open on the right: drop out, administrative censoring

Open on the left: late entry

# Learn from non-closed cohorts



# Learn from non-closed cohorts



# Cohorts can be open on the right even without dropout!

Say you wish to estimate the 5-year risk of death among people entering HIV care.

You enroll people in the study at entry into care (origin) between 2012 and 2020.

If your methods require a closed cohort, you require 5 full years of follow-up, so can only use data on those enrolled between 2012 and 2015, essentially throwing away 5 years of data.

But using time-to-event methods, the partially observed person time (and any events) among those enrolled between 2015 and 2020 also contribute to the study.

# Consider time-varying exposures

- Nonadherence
- Policy changes

Critical to compare the right people at the right times!

# Chapter 4: Line diagrams

# Line diagrams

Visualize choices of origin and timescale

See which participants are being compared at a given point in time.

In next week's reading, we will see how choice of timescale results in very different line diagrams and very different risk functions.

# Line diagrams, steps

1. Draw and label axes (y axis is usually studyid, x axis is timescale, starting at the origin)
2. Calculate the amount of time after the origin that the first person enters the study. Place an open circle at this timepoint.
3. Draw a line from this open circle until the last available information for that participant.
4. If an event was observed, place a closed circle at that time
5. If no event was observed, place an arrow at that time
6. Repeat for remaining participants

# An example, drawing line diagrams.

Say have a (small) cohort study of 5 soldiers selected at random from all people enlisting in the armed services between 2004 and 2014 and followed for mortality up to 10 years.

Complete line diagrams as described by the following 3 slides.

# Example data

Draw a line diagram with the  $x$  axis as *age (starting at age 18)* and the  $y$  axis as soldier id number.

ID number	Age at enlistment	Enlistment date	Last date with available information	Vital status at last info
1	24	1 July 2008	Today()	Alive
2	18	1 July 2008	1 January 2011	Dead
3	21	1 January 2011	15 June 2016	Dead
4	35	1 January 2011	Today()	Alive
5	19	1 July 2013	Today()	Alive

Recall: Say we have a (small) cohort study of 5 soldiers selected at random from all people 18 and over enlisting in the armed services **between 2004 and 2014** and followed for mortality **up to 10 years**.

# Example data

Draw a line diagram with the *x* axis as *calendar time (starting 1 Jan 08)* and the *y* axis as soldier id number.

ID number	Age at enlistment	Enlistment date	Last date with available information	Vital status at last info
1	24	1 July 2008	Today()	Alive
2	18	1 July 2008	1 January 2011	Dead
3	21	1 January 2011	15 June 2016	Dead
4	35	1 January 2011	Today()	Alive
5	19	1 July 2013	Today()	Alive

Recall: Say we have a (small) cohort study of 5 soldiers selected at random from all people 18 and over enlisting in the armed services **between 2004 and 2014** and followed for mortality **up to 10 years**.

# Example data

Draw a line diagram with the  $x$  axis as *time since enlistment* and the  $y$  axis as soldier id number.

ID number	Age at enlistment	Enlistment date	Last date with available information	Vital status at last info
1	24	1 July 2008	Today()	Alive
2	18	1 July 2008	1 January 2011	Dead
3	21	1 January 2011	15 June 2016	Dead
4	35	1 January 2011	Today()	Alive
5	19	1 July 2013	Today()	Alive

Recall: Say we have a (small) cohort study of 5 soldiers selected at random from all people 18 and over enlisting in the armed services **between 2004 and 2014** and followed for mortality **up to 10 years**.

# L1: Questions and time in epidemiologic studies

EPID 722

Spring 2021

UNC – Chapel Hill

jessedwards@unc.edu