Public Health Sciences 310
Epidemiologic Methods

Lecture 3
Measuring Disease Occurrence

January 11, 2024

Brian C.-H. Chiu

Measures of Disease Occurrence

Epidemiology Setting

- **Incidence: # new** cases during a specified period of time in a population at risk for developing the disease.

  * Cohort Study *

  ☆Causal Association

Public Health Setting

- **Prevalence**: the amount of disease already present in a population

Incidence

Undiseased → Diseased

- Incidence is a measure of transitions and thus of risk (the probability that an event will occur); it must have a specified period of time. ex) Developing sepsis from burned patients between 6 month v.s 1 year (risk very different)!
- An individual included in the denominator must have the potential to become part of the group that is in the numerator.
  At risk for developing disease
  - Do not have condition already
  - Not ineligible ex) Analyze men in IL with prostate cancer. Denomenator can't include women - not eligible.

Strategies for Calculating Incidence

Two major ways to define incidence

- Chance **Cumulative incidence (risk)**: the likelihood that an individual will contract a disease
  - Survival analysis over time
  - Unit of analysis = individual (i.e., **persons at risk**)
- Rate **Incidence Rate or Density**: how fast new occurrence of disease/event arise
  - Analysis based on person-time
  - Unit of analysis = time (i.e., Rate **person-time units at risk**)

Cumulative Incidence (or Risk)

Cumulative Incidence:

$$\frac{\text{number of new cases of disease}}{\text{population at risk}} \times \text{unit multiplier per specified time interval}$$

a) all in denominator must be at risk for developing the disease
b) all cases in numerator must come from the denominator
c) fixed cohort followed for complete time interval v.s. open cohort (allow new/old participants to enter/leave)
d) measure of the risk or probability of developing disease in time interval
e) not affected by natural history of illness or treatment
f) useful for studying disease etiology Prognostic factor in clinical study

Example

- Follow up of 10 patients for 2 years
  - 6 Deaths
  - 3 censored before 2 full years of follow-up
  - 1 survived for 2 full years

- Question: What is the Cumulative Incidence (or the Cumulative Survival) up to 2 years? $\boxed{\frac{6}{10} = 0.6}$, or 60% risk of developing colon cancer during 2 year follow-up.

$\frac{\#\text{ new cases}}{\#\text{ persons}}$

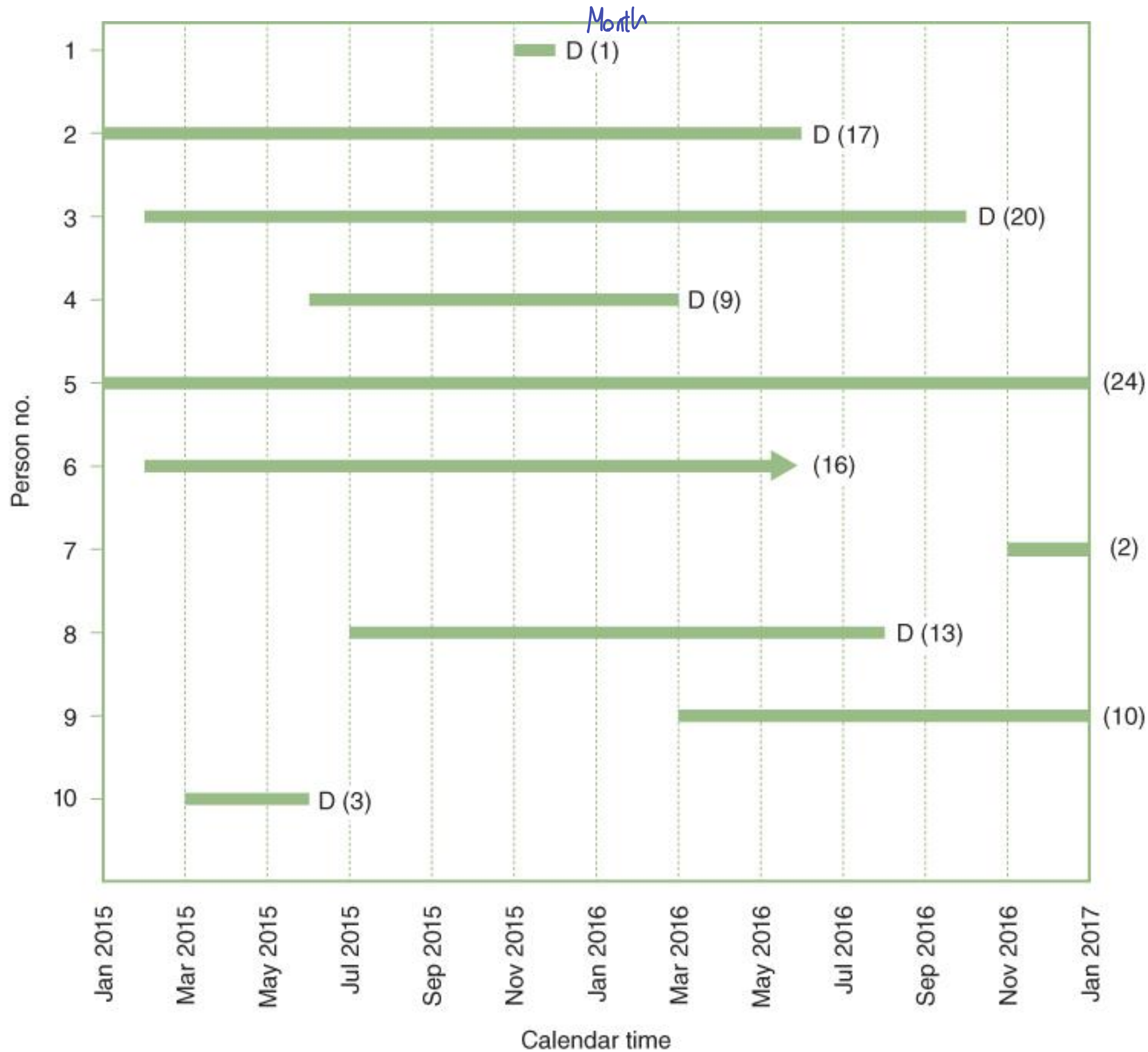
Figure 2.1: Hypothetical cohort of 10 persons followed for up to 24 months from January 2015 through December 2016. D, death; arrow, censored observation; ( ), duration of follow-up in months (all assumed to be exact whole numbers).

Death

Censored observation (lost to follow-up, withdrawal), migration... etc

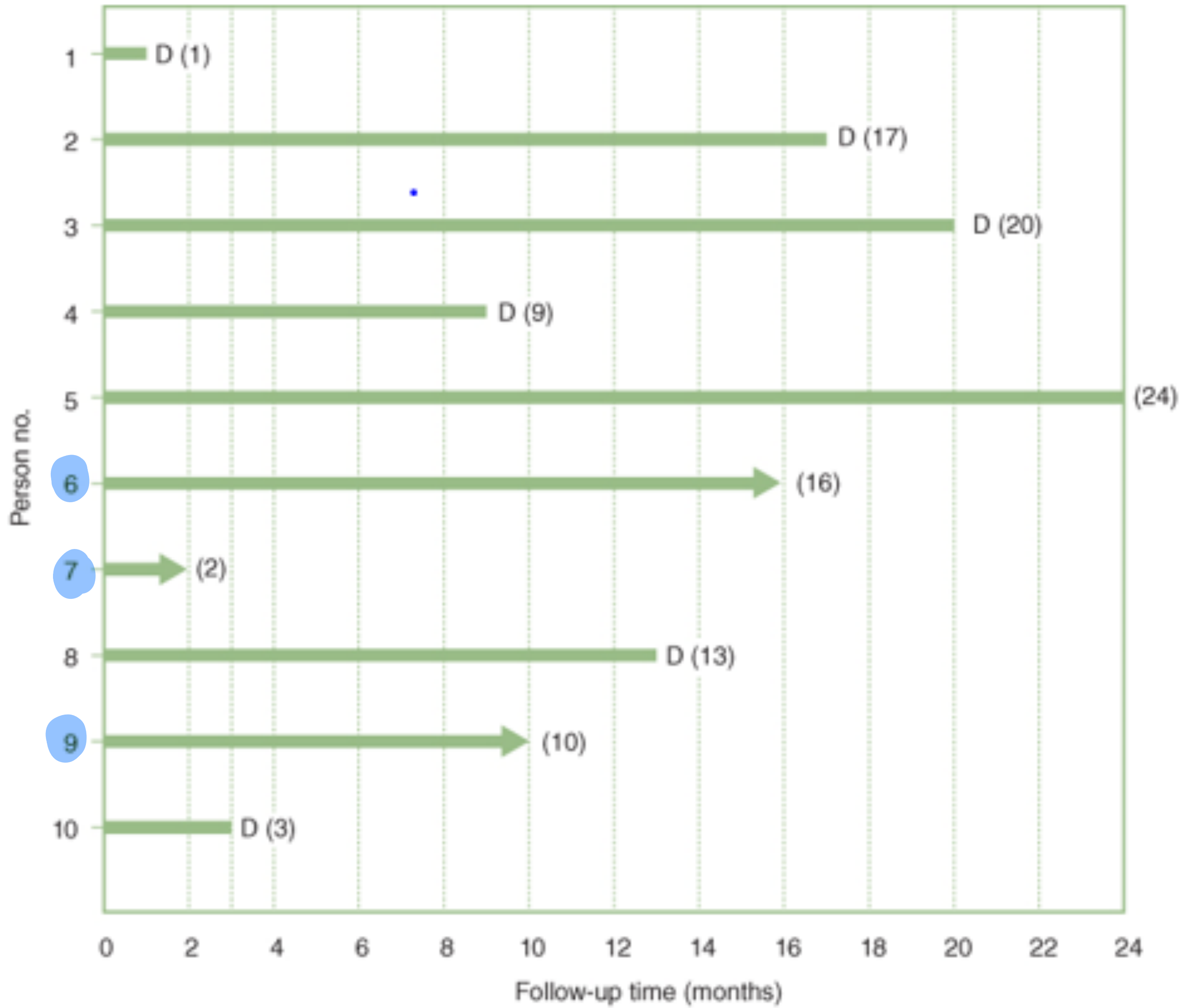
Administrative (finish the study)

Number of months of follow-up (#)



Change time scale to "follow-up" time: compare length/duration

Figure 2.2: Same cohort as in Figure 1, with person-time represented according to time since the beginning of the study. D, death; arrow, censored observation; ( ), duration of follow-up in months (all assumed to be exact whole numbers).



Calculating Cumulative Incidence and Survival

- Cumulative Incidence (up to time "t")

$$q_t = \frac{Number\ of\ individuals\ with\ the\ event\ by\ t}{Number\ at\ risk\ at\ baseline}$$

- Cumulative Survival

$$S(t) = \boxed{1 - q_t} = \frac{Number\ of\ individuals\ alive\ beyond\ t}{Number\ at\ risk\ at\ baseline}$$

Example

- Assume NO censoring
- The two years cumulative incidence (cumulative probability of disease/event) = 6/10=0.6 (or 60%)

  Death
- The two years cumulative survival = ? $\boxed{1-0.6=0.4, \text{ or } 40\%}$

  Alive

  $1 - \text{incidence} = 1-p=q$

Problem: requires accounting for censoring (losses to follow-up)

3

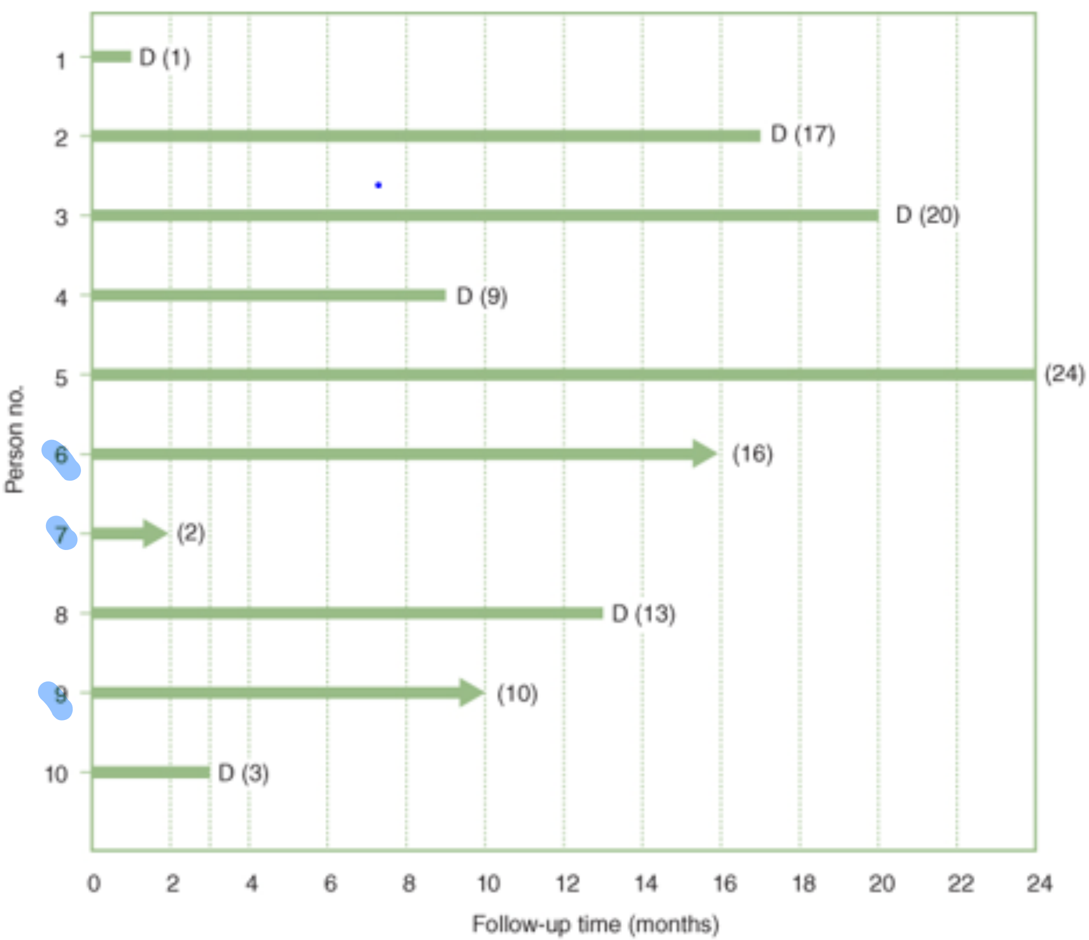
* Approach to Accounting Censoring

One solution: Classic Life Table

(Short follow-up only)

- Assume that censored observations over the period contribute one-half the persons at risk in the denominator (censored observations occur uniformly throughout the follow-up interval).

Assume constant censoring risks



$$q_{2\ years} = \frac{6}{10 - \boxed{{}^{1}/_{2} \times 3}} = \frac{6}{8.5} = 0.71$$

# of censoring individuals

$$S(2\ years) = 1 - q_{2\ years} = 0.29$$

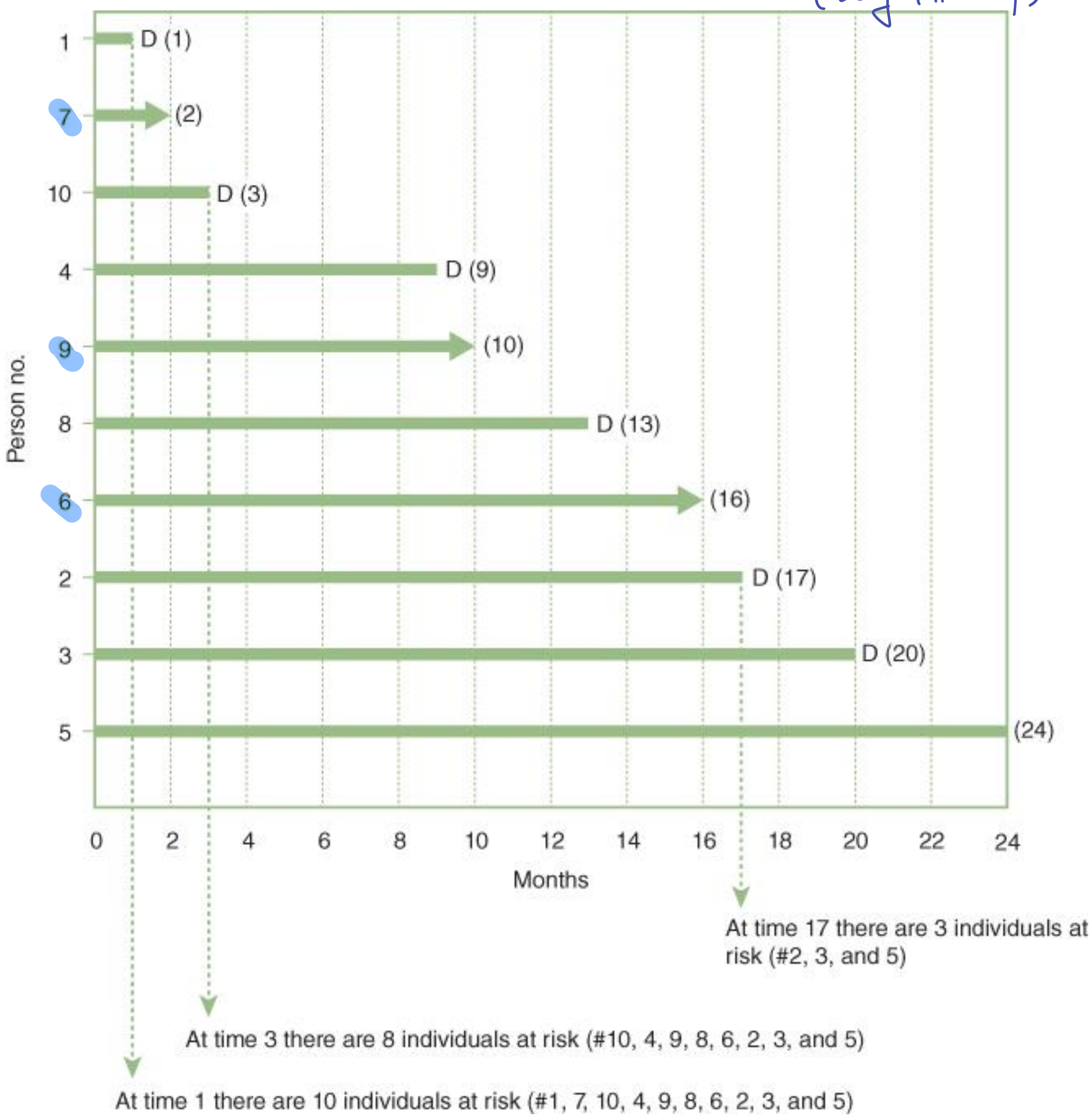
Another problem

- If the follow-up is long, the risks cannot be assumed to be constant, and thus, the follow-up time needs to be partitioned.

* Approach to Accounting Censoring

Solution: Kaplan-Meier Approach

(Long follow-up)

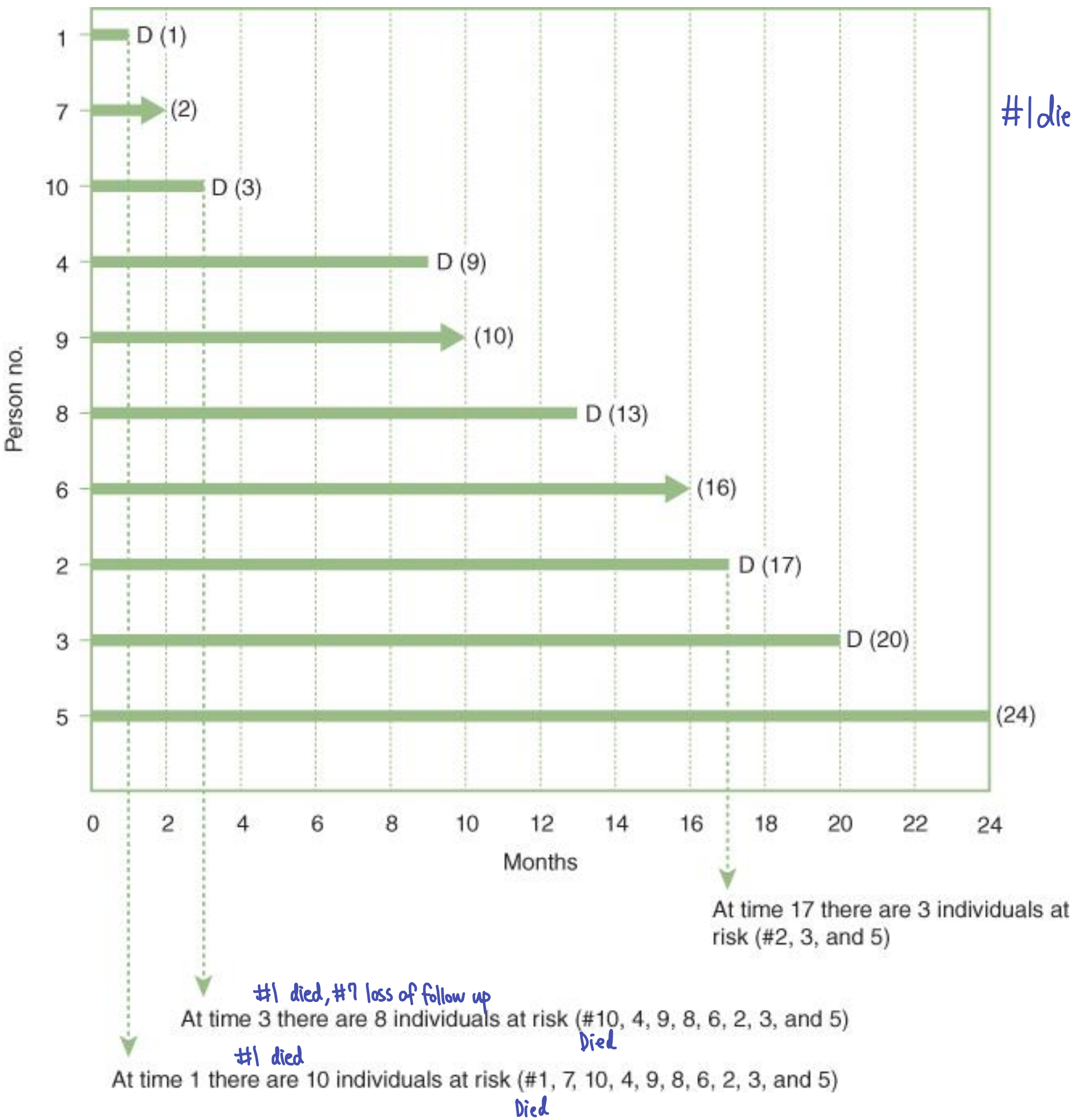


- Calculate the cumulative probability of event (and survival) based on conditional probabilities at each event time.
- Step 1: Sort the follow-up time from shortest to longest

Figure 2.3: Same cohort as in Figures 1 and 2, with individuals sorted according to follow-up time from shortest to longest. D, death; arrow, censored observation; ( ), duration of follow-up in months (all assumed to be exact whole numbers). As examples, the vertical arrows mark the individuals who were at risk for the calculations of the conditional probabilities of death at three of the event times: 1 month, 3 months, and 17 months.

- Step 2: For each time of occurrence of an event, compute the conditional survival (What is the survival proportion beyond each event?)

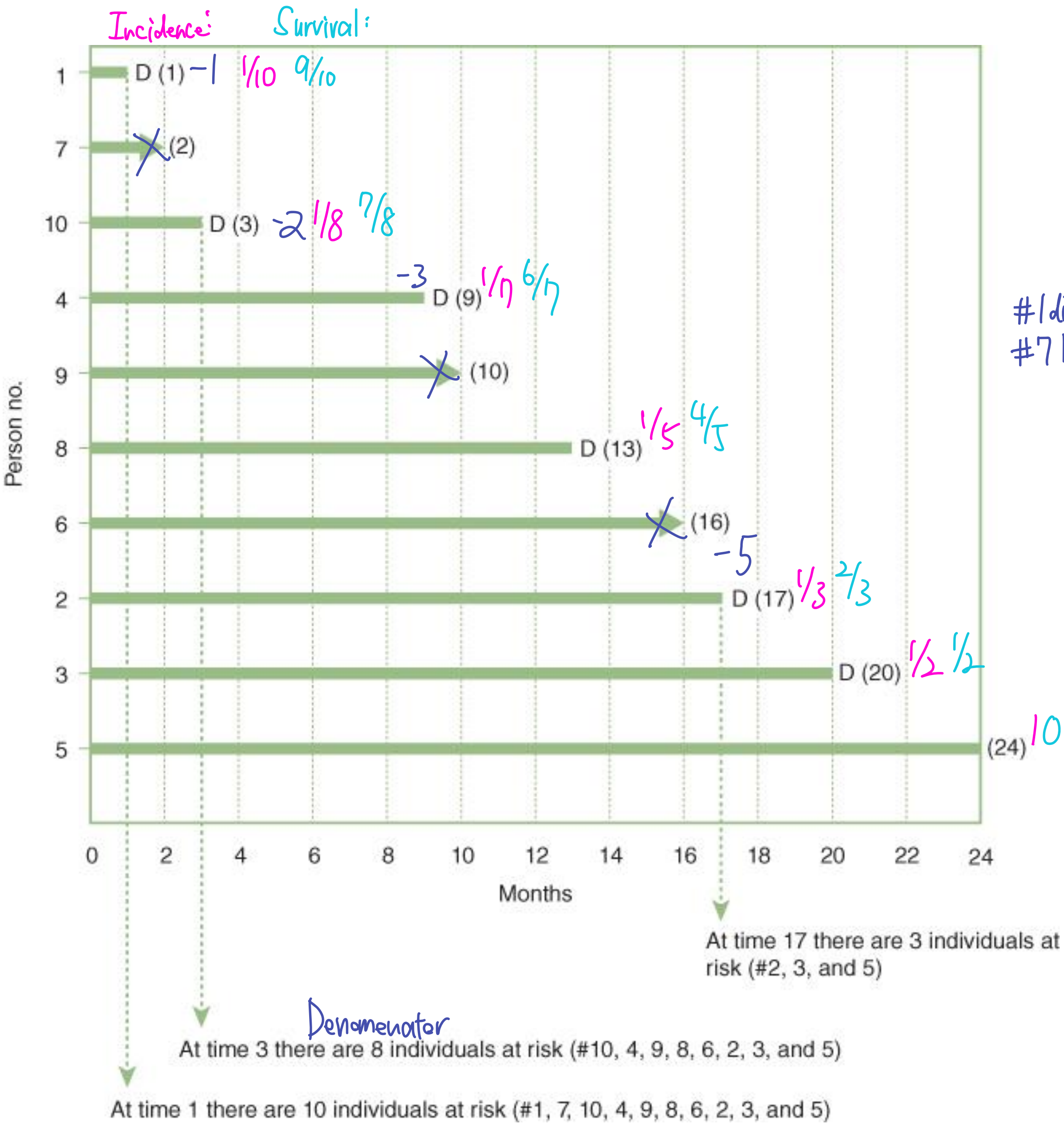
recalculate each time during new event: (death) of individual



When the first event occurs #1 died (1 month after beginning of follow-up), there are 10 Besides #1. persons at risk. One dies at that point; 9 of the 10 survive beyond that point. Thus:

-Incidence of event at exact time 1 month: 1/10=0.1

-Probability of survival beyond 1 month: 9/10=0.9



#1 died, #7 loss.

When the <u>second event</u> occurs (3 months after beginning of follow-up), there are 8 persons at risk. One dies at that point; 7 of the 8 survive beyond that point. Thus:

-Incidence of event at exact time 3 months: 1/8=0.125

Not 2 b/c #7 does not count for loss of follow-up

-Probability of survival beyond 3 months: 7/8=0.875

Conditional Probability of an Event (or of survival)

- The probability of an event (or of survival) at time t (for the individuals at risk at time t), this is, conditioned on being at risk at exact time t.
  excluding loss of follow up

- Step 3: For each time of occurrence of an event, compute the cumulative survival (survival function), multiplying conditional probabilities of survival.

Step 3

- Cumulative survival: $$S(t_i) = \prod_{\substack{All\ deaths \\ t_j < t_i}} \left(1 - \frac{d_j}{n_j}\right)$$

TABLE 2-3 Calculation of Kaplan-Meier Survival Estimates for the Example in Figure 2-3

Time (Months) (1) t	Individuals at Risk (2) n_i	Number of Events (3) d_i	Conditional Probability of the Event (4) $q_i = d_i/n_i$	Conditional Probability of Survival (5) $p_i = 1 - q_i$	Cumulative Probability of Survival* (6) S_i	
1	10†	1	1/10 = 0.100	9/10 = 0.900	0.900 90%	
3	8†	1	1/8 = 0.125	7/8 = 0.875	0.788 79%	= 0.900 × 0.875
9	7	1	1/7 = 0.143	6/7 = 0.857	0.675 68%	= 0.788 × 0.857
13	5	1	1/5 = 0.200	4/5 = 0.800	0.540 54%	⋮
17	3†	1	1/3 = 0.333	2/3 = 0.667	0.360 36%	
20	2	1	1/2 = 0.500	1/2 = 0.500	0.180 18%	

*Obtained by multiplying the conditional probabilities in column (5)—see text.
†Examples of how to determine how many individuals were at risk at three of the event times (1, 3, and 17) are shown with vertical arrows in Figure 2-3.

Plotting the survival function:

Remember to 1-p=q

Cumulative Probability of Survival* (6) S_i	
0.900	$1-\frac{1}{9}$
0.788	$1-\frac{1}{8}$
0.675	⋮
0.540	
0.360	
0.180	

Don't plot loss of follow-up.

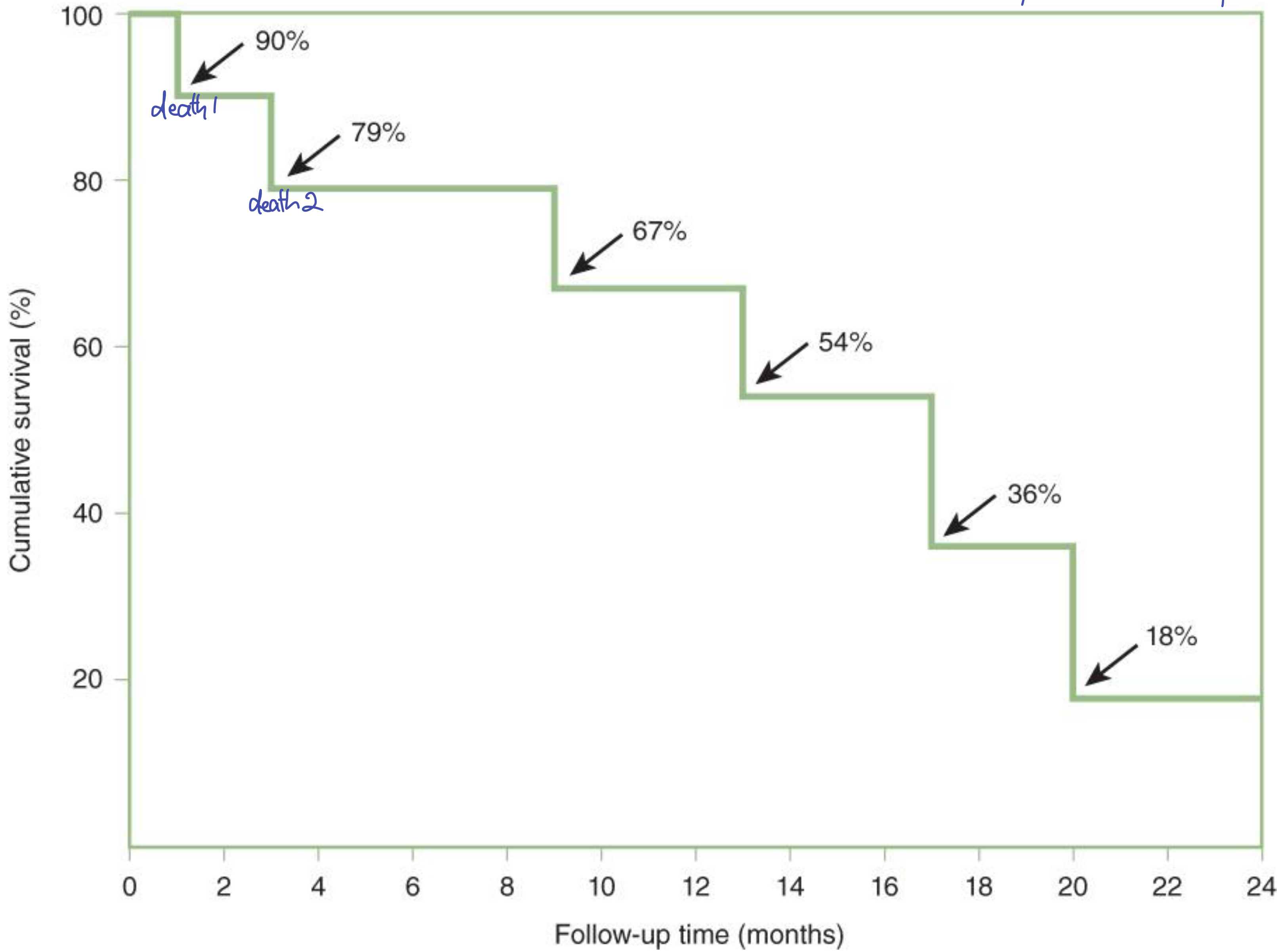


Figure 2.4: Kaplan–Meier curve corresponding to the data in Table 2.3, column 6

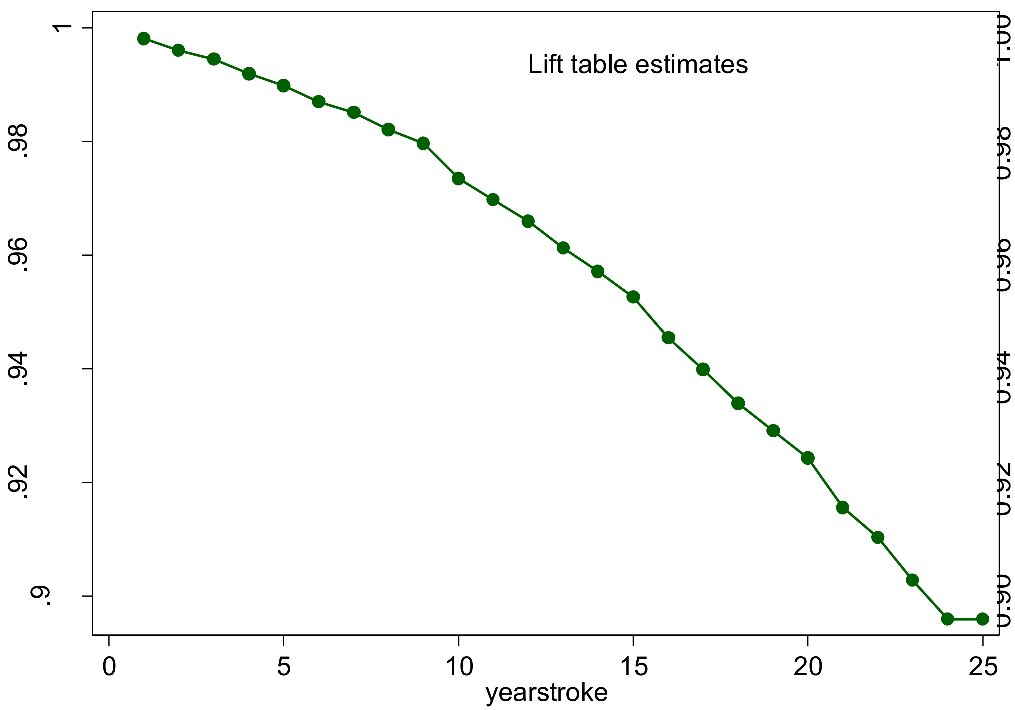
Assumptions in Survival Analysis

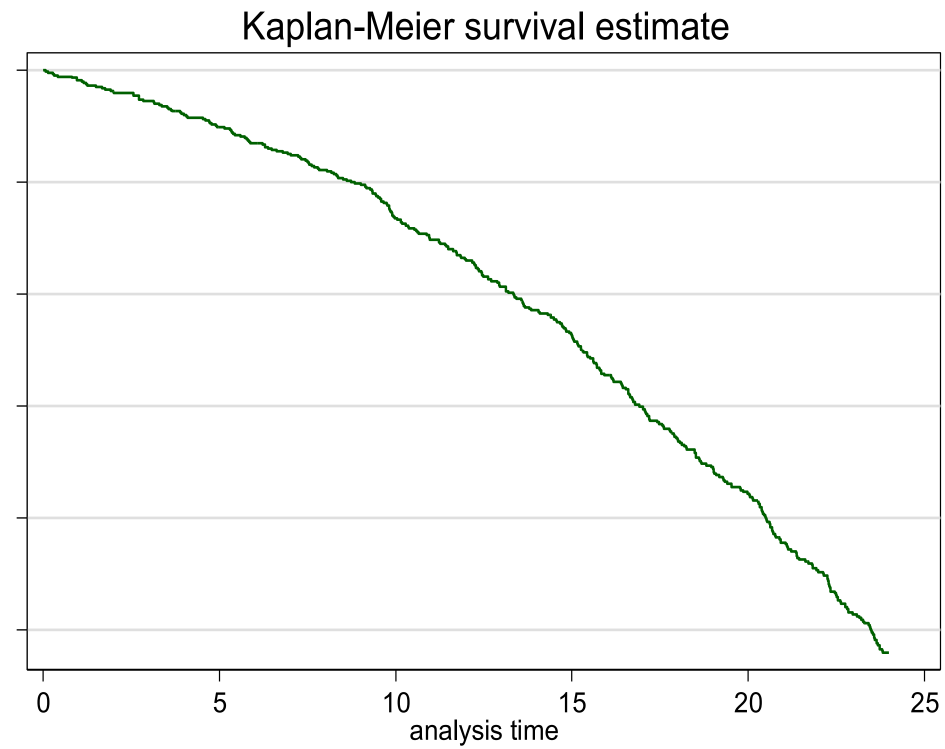
- **Censoring is independent of survival**
  - Uninformative censoring: those censored at time t have the same prognosis as those remaining. V.S. Informative censoring
  - Types of censoring:
    - Lost to follow-up (e.g., migration, refusal)
    - Death (from another cause)
    - Administrative withdrawal (study finished)

Summary of Life Table vs. Kaplan-Meier

n >1000 Few months

- With large samples (N), or short intervals there should not be much difference

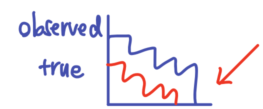




Summary of Life Table vs. Kaplan-Meier

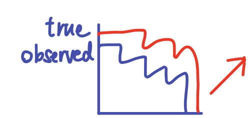
- With large samples (N), or short intervals there should not be much difference
- A life-table assumes that incidence is constant within each interval
- Both assume no secular 長期 trend exists Prolonged period (years)
- Both assume censoring is independent of survival
  - If censored true observations tend to have worse prognosis than those remaining in the study: Overestimate

    Observed survival > True survival



  - If censored true observations tend to have better prognosis than those remaining in the study: Underestimate

    Observed survival < True survival



Strategies for Calculating Incidence

Two major ways to define incidence

- **Cumulative incidence (risk)**: the likelihood that an individual will contract a disease
  - Survival analysis
  - Unit of analysis = individual (i.e., **persons at risk**)
- **Incidence Rate or Density**: how fast new occurrence of disease/event arise
  - Analysis based on person-time
  - Unit of analysis = time (i.e., **person-time units at risk**)

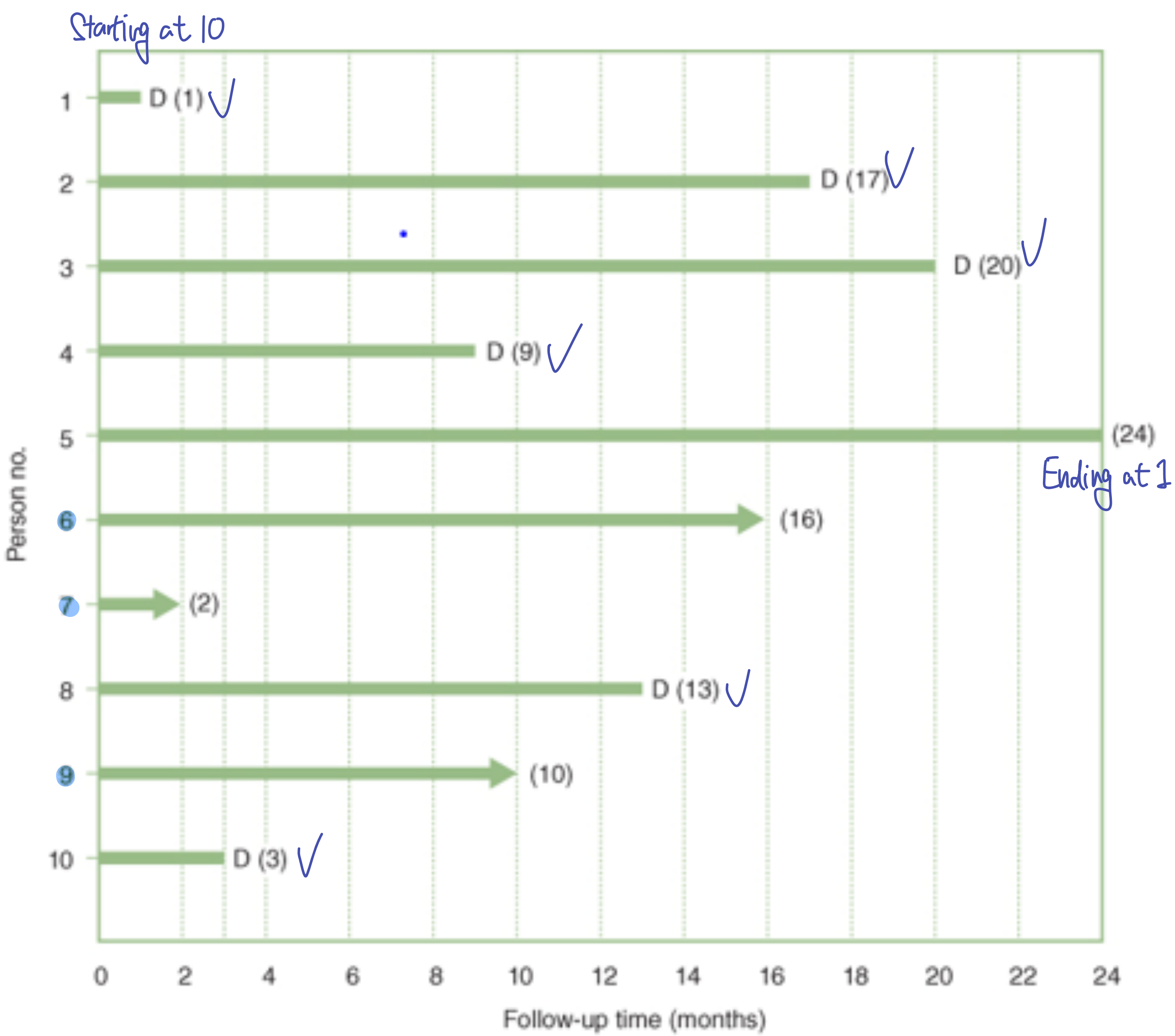
Incidence (or Mortality) Rate: Aggregate Data

Group of study participants

- Denominator is average population during period.
- Period of time short
- Assumes population is stable

$$\text{Incidence rate} = \frac{\text{No. of new events during period}}{\text{Average population at-risk during period}}$$

Incidence (or Mortality) Rate: Aggregate Data



Average population (at risk)

n = (10+1)/2 = 5.5

start end

Or, can also be calculated by subtracting one-half of the events (D) and losses (C) from the initial population:

n = 10 – ½ (6+3) = 5.5

loss of follow-up

Not #5 b/c still alive

Incidence rate ex) Health Policy

events

= 6/5.5

=1.09 per person-2yrs

=0.545 per person-year

$\frac{1.09 \text{ person} \times 1/2}{2 \text{ years} \times 1/2} = \frac{0.545 \text{ person}}{\text{Year}}$

(Most Accurate)

Incidence (or Mortality) Density: Individual Data

- Used for data when precise timing of events/censoring are known or can be reasonably estimated.
- Denominator is **person-time**.
- Assumes censoring and survival are independent

$$\text{Incidence density} = \frac{\text{No. of new events during period}}{\text{Total person-time at risk}}$$

Person-Time Example

- Step 1: Calculate denominator, i.e. units of time contributed by each individuals, and total:

TABLE 6 Calculation of the number of person-years based on Figure 2.

Person no.	Total follow-up (in months)	Contribution to the total number of person-years by participants in:		
		1st Year of follow-up	2nd Year of follow-up	Total follow-up period (Sum of 1st + 2nd Year)
1	1 Died within a year...	1/12 = 0.083	0 Subject died	0.083
2	17 Survived the first year...	12/12 = 1.000 Subject survived	5/12 = 0.417	1.417
3	20	12/12 = 1.000	8/12 = 0.667	1.667
4	9	9/12 = 0.750	0	0.750
5	24	12/12 = 1.000	12/12 = 1.000	2.000 survived
6	16	12/12 = 1.000	4/12 = 0.333	1.333 censor
7	2	2/12 = 0.167	0	0.167 censor
8	13	12/12 = 1.000	1/12 = 0.083	1.083
9	10	10/12 = 0.833	0	0.833 censor
10	3	3/12 = 0.250	0	0.250
Total	115 months	7.083 years	2.500 years	9.583 years

Person-Time Example

- Step 2: Calculate rate per person-year for the total follow-up period:

$$\boxed{Incidence\ Density = \frac{Number\ of\ events}{Total\ person\ time}}$$

From previous table:

√

$$R = \frac{6}{9.583\ years} = 0.63\ per\ person\ year$$ V.S. 0.545

Comparing Measures of Incidence

EXHIBIT 1 Comparing measures of incidence: cumulative incidence vs incidence rate.

	Cumulative incidence Chance		Incidence density/rate Rate	
	If follow-up is complete	*If follow-up is incomplete*	*Individual data (cohort)*	Aggregated *Grouped data (area)*
Numerator	Number of cases	Classic life table Kaplan–Meier	✓ Number of cases	✓ Number of cases
Denominator	Initial population		✓ sum of each Person-time Calculate for each then add	✓ Average population* one time calculation
Units	Unitless		✓ Time^{-1} Rate	
Range	0 to 1		0 to infinity	
Synonyms	Proportion chance Probability		Incidence density†	

*Equivalent to person-time when events and losses (or additions) are homogeneously distributed over the time interval of interest.

†In the text, the term *density* is used to refer to the situation in which the exact follow-up time for each individual is available; in real life, however, the terms *rate* and *density* are often used interchangeably.

Prevalence

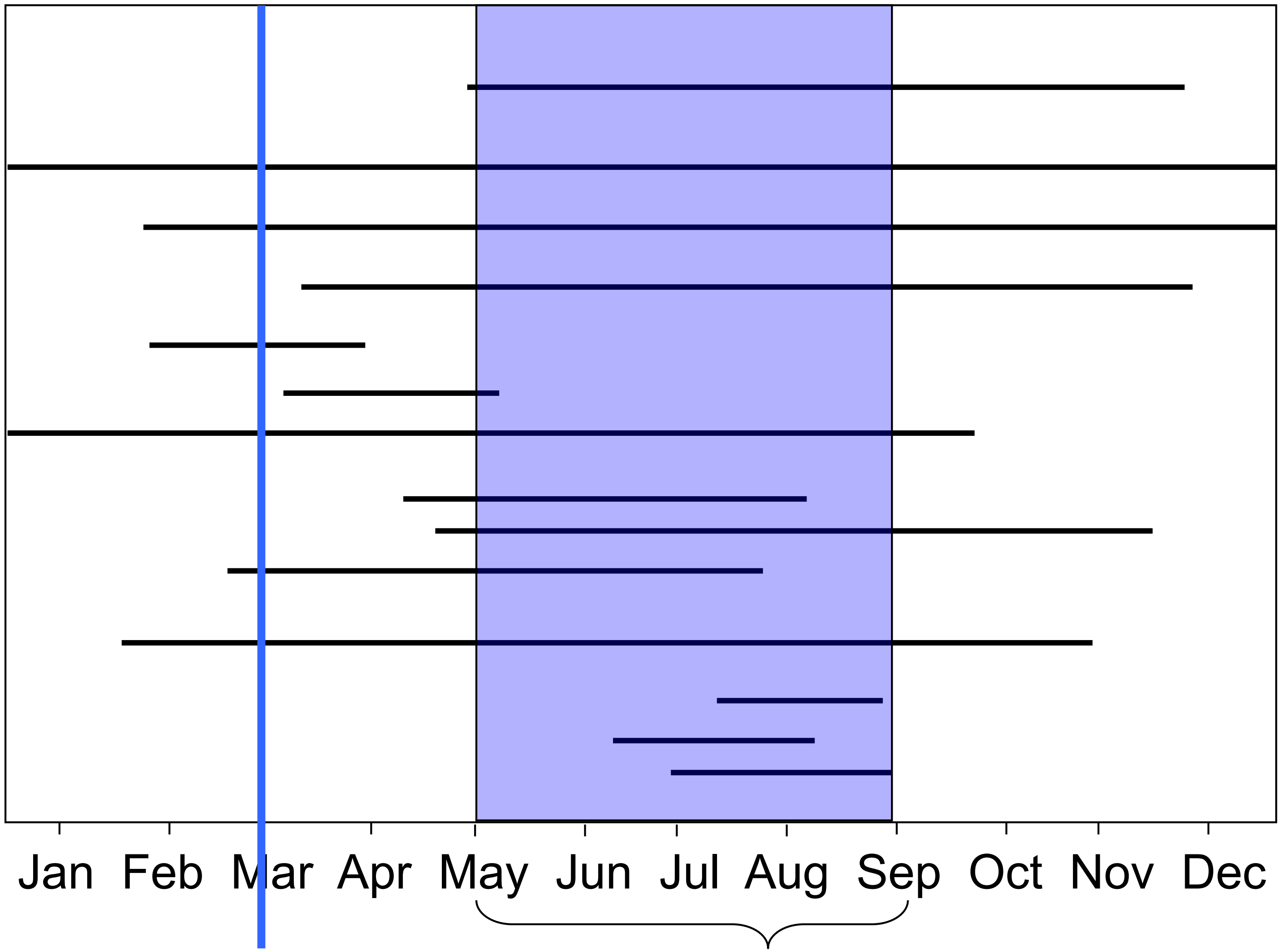
- The frequency (proportion) of people with outcome

- Point prevalence is ascertained at a specific point in time.

- Period prevalence includes all new and existing cases within a defined time frame.

Prevalence

Hypothetical study of persons with symptoms of avian Influenza in a city of 1 million people.



Point prevalence
6 cases / million pop.

Period prevalence (4 mo.)
13 cases / million pop.

Prevalence and Incidence

- Prevalence is a slice through the population at one time and does not take into account duration.
- Some persons will have had the disease for 20 years; some will be relatively new cases. Thus it is not a measure of risk.
- Prevalence is the most helpful measure of the *burden of disease on the population.* "Impact"
- Incidence is the most helpful measure of *risk of disease in the population.* "Risk"
- Prevalence = incidence x duration
  - When incidence and duration are stable

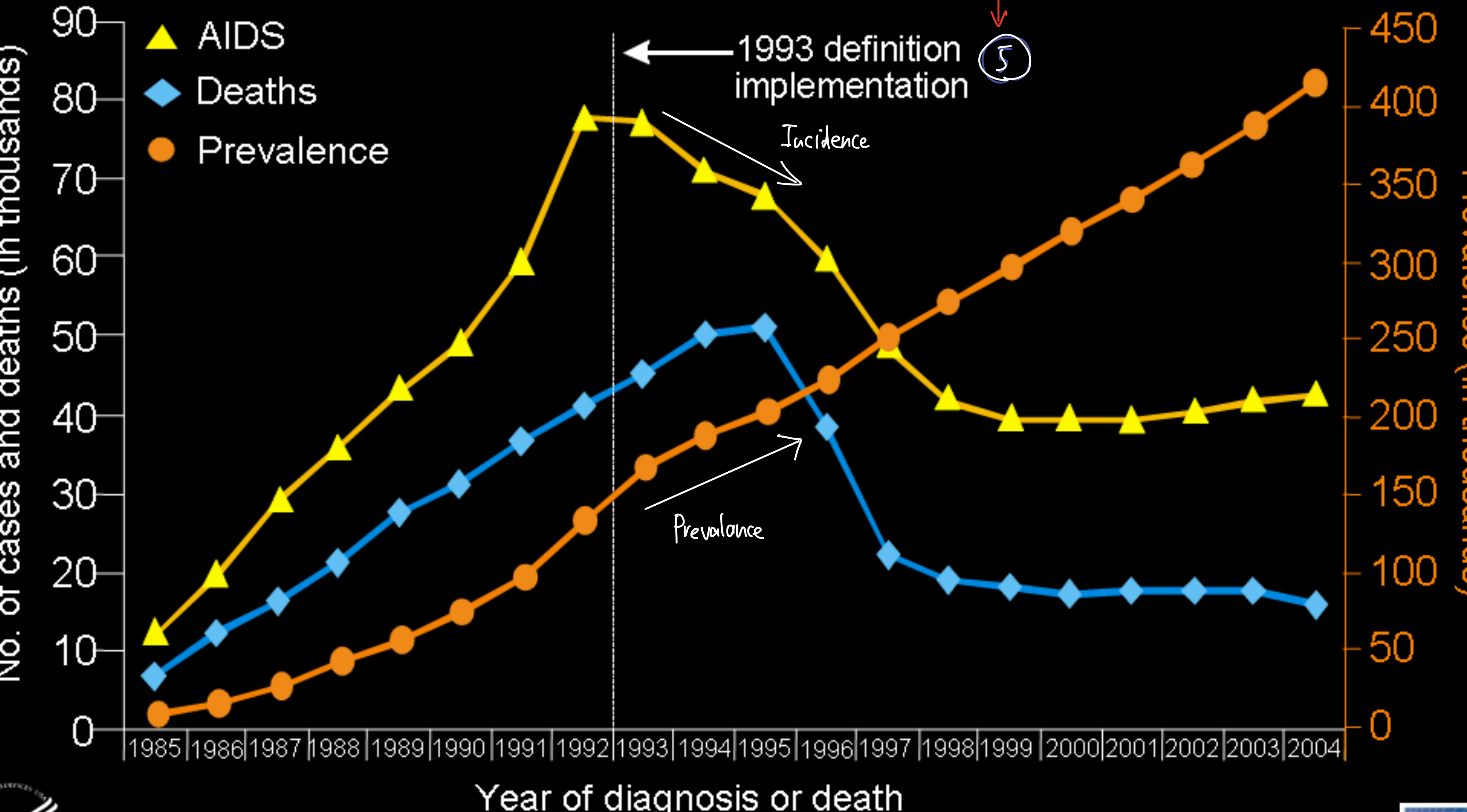
- If incidence is stable, what would cause prevalence to increase? Examples?

Total # of cases in population

1. ↑ Survival: Improved medical treatments, better healthcare to enhance survival.
2. ↓ Mortality: Condition mortality ↓.
3. Population Growth
4. Changes in risk factor

⑤ Improved diagnosis and detection: New technology developed to change diagnosis

AIDS Cases, Deaths, and Persons Living with AIDS, 1985-2004, United States



Note. Data have been adjusted for reporting delays.