UP-ART Training
Research Methods Part 3: Introduction to Statistical Analysis II

Lizzie Chappell
Siobhan Crichton
Edith Milanzi

INSTITUTE OF CLINICAL TRIALS AND METHODOLOGY

Aim of session

To be able to…

- Understand the difference between incidence and prevalence
- Know how to interpret a risk difference, risk ratio and odds ratio
- Interpret results from different types of survival analysis: a Kaplan Meier graph, a log-rank test and a hazard ratio
- Understand the difference between adjusted and unadjusted estimates from a regression analysis

Outline

1. Recap on confidence intervals and p-values
2. Incidence and prevalence
3. Outcome measures for binary data (risk difference, risk ratio, odds ratio)
4. Survival analysis
5. Regression

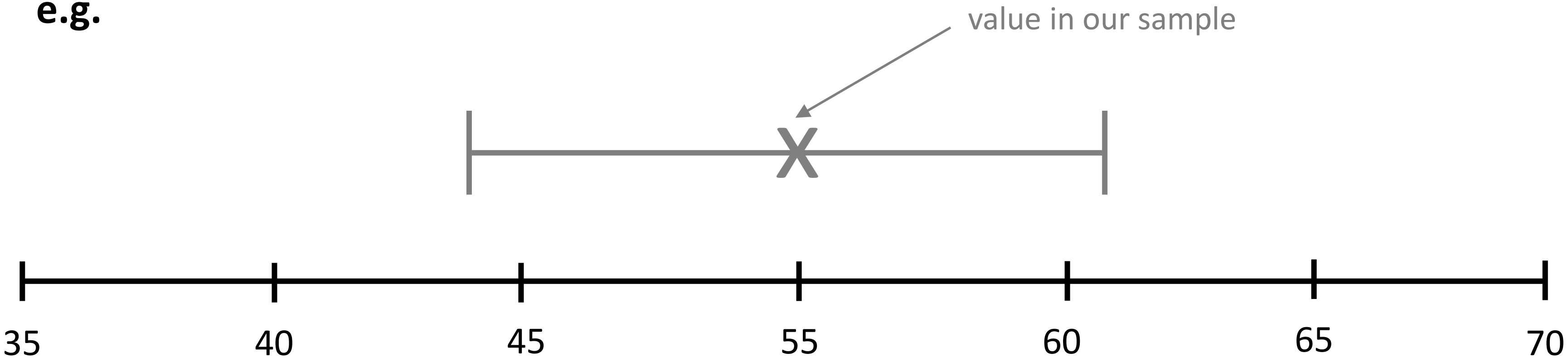
Recap on different types of outcome measure

Type of data	Summary measures	To make comparisons
Continuous	Means	Difference in means
Binary	Proportions	Compare proportions (e.g. risk difference, risk ratio), odds ratio
Survival time	Kaplan Meier plots, survival time	Log rank test, hazard ratio

Recap of confidence intervals

confidence interval = range of values we are reasonable sure contains the true population value

e.g.



"We can be 95% sure that the true mean weight is between 44.2 and 61.3 kg"

Recap of p-values/hypothesis testing

p-value = probability of observing the results we have observed if H_0 were true

e.g.

H_0: mean weight for those on treatment A = mean weight for those on treatment B
H_1: mean weight for those on treatment A ≠ mean weight for those on treatment B

Mean in group A = 47.0 kg
Mean in group B = 42.3 kg
Difference = 4.7 kg

→ Conduct a hypothesis test → Calculate p-value = 0.02

→ "If the two treatments were actually identical, there is a 2% chance we would see a difference as large as 4.7 kg by chance"

→ Seems unlikely… so conclude there is evidence of a difference

Recap of link between confidence intervals and p-values

95% CI includes value being tested in H_0	⟷	P-value > 0.05 Not enough evidence to reject H_0
95% CI does not include value being tested in H_0	⟷	P-value < 0.05 H_0 rejected

Example: blood pressure

Equivalent

CI = -4.6, 10.6	→	includes 0 mmHg
p-value = 0.403	→	cannot reject H_0

INCIDENCE/PREVALENCE

Prevalence and incidence

- Prevalence is the number of cases in the population at a given time point
  - E.g. 5% of children within Uganda are living with HIV

- Incidence is the number of new cases in a population over a specific time period
  - Often expressed per 1000, 10,000, 100,000, .. population
  - Sometimes expressed in terms of *person years*
  - E.g. the incidence of HIV in children in Uganda in 2017 was 5.6 cases per 1000 children

- As timing of outcomes/events is known in a cohort, they can be used to measure incidence

Formulae

Prevalence

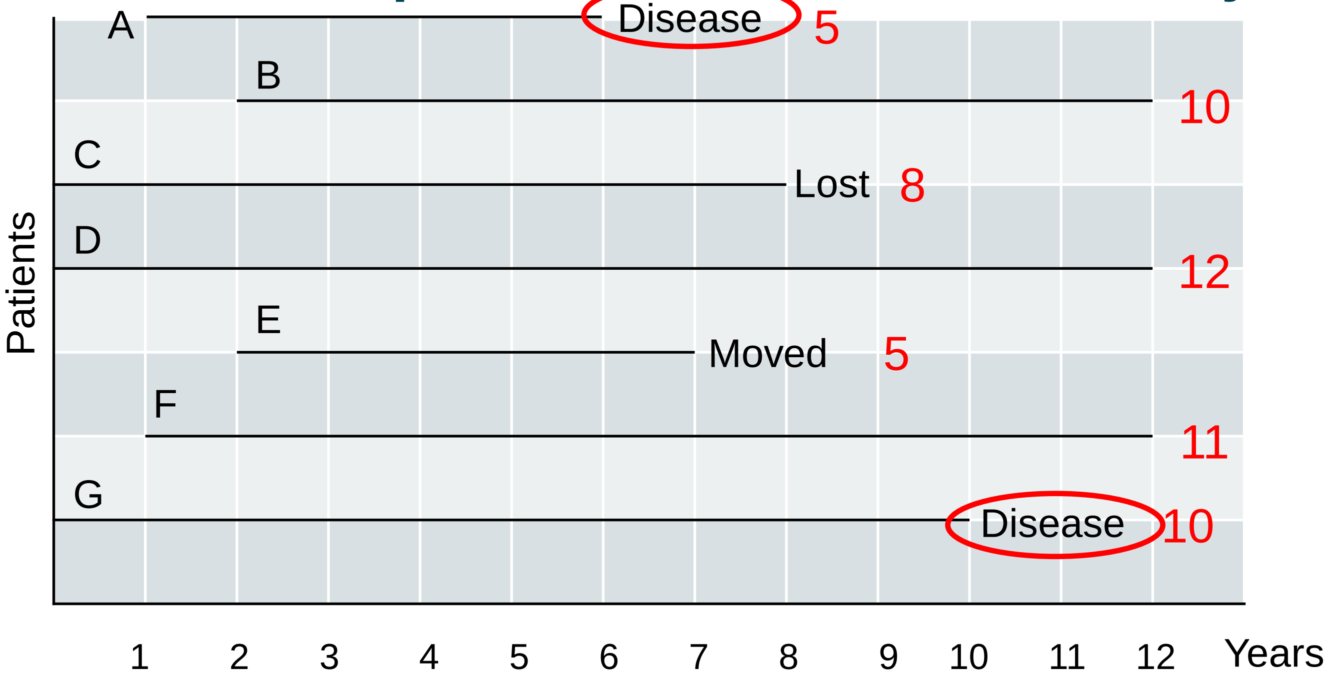
Number of cases or events in a population/ Total population

Incidence rates

$$\frac{\text{Number of new cases or events in a given time period}}{\text{Sum of the length of time during which each person in the population is at risk}}$$

(number of persons x time contribution)
(person time, usually measured in years)

Schematic representation of a cohort study



2 cases of disease / (5+10+8+12+5+11+10 person years) = 2 / 61 = 0.033 per person year
= 3.3 per 100 person years

Figure adapted from J Giesecke. Modern Infectious Disease Epidemiology. 1994, Edward Arnold, UK,

Example

RESEARCH Open Access

Prevalence and incidence rate of tuberculosis among HIV-infected patients enrolled in HIV care, treatment, and support program in mainland Tanzania

- Retrospective cohort-Enrolled HIV clients in HIV care/treatment between January 2011 and December 2014 in Tanzania.
- 527, 249 individuals with a total of 11,539,844 clinical encounters enrolled
- Aimed to assess the prevalence and TB incidence rate per 1000 person-years.

Majigo et al, Trop Med Health 48, 76 (2020)

Example

Table 1 Number of individuals and clinic encounters in HIV care, treatment, and support program in Tanzania from 2011 to 2014

Program year	Number of individuals	Total clinic encounters
2011	427,117	2,560,290
2012	449,114	2,565,557
2013	461,857	3,004,427
2014	527,249	3,409,570
Total		**11,539,844**

Table 2 Prevalence of TB among individuals enrolled in HIV care, treatment, and support program in Tanzania from 2011 to 2014

Characteristics	2011	2012	2013	2014	Total
	N (%)	***N*** (%)	***N*** (%)	***N*** (%)	***N*** (%)
Overall	8765 (2.1)	9798 (2.3)	11,212 (2.5)	9857 (1.9)	39,632 (2.2)

Table 3 Incidence of tuberculosis among individuals enrolled in HIV care, treatment, and support program in Tanzania in 2011 to 2014

Variable	TB cases	1000 person-years	TB incident rate/1000 person-years (95%CI)	
Overall	22,071	1323.6	16.7 (16.4–16.9)	-

Majigo et al, Trop Med Health 48, 76 (2020)

Can also stratify by key variables…

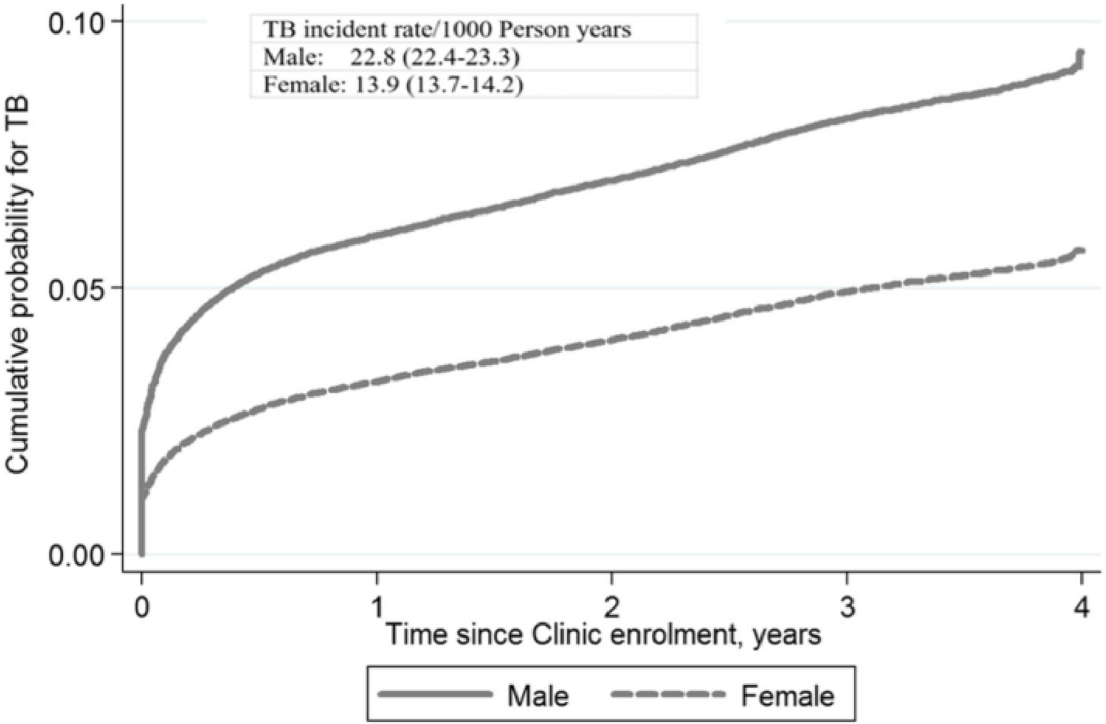


Fig. 3 Cumulative probability of TB incident after enrollment to HIV care and treatment by age and sex

Majigo et al, Trop Med Health 48, 76 (2020)

OUTCOME MEASURES FOR BINARY DATA

HIV prevention trial

- HIV negative participants are recruited
- Randomised to receive pre-exposure prophylaxis (PrEP) or no PrEP
- Objective: to identify whether PrEP is effective in preventing HIV
- Participants are followed for 6 months and are tested for HIV at each visit

- **Binary outcome** - HIV-positive *or* HIV-negative 6 months after randomisation

The 2x2 table

- Two arms of the study (here PrEP and No PrEP)

- We can use a 2x2 table to display our binary outcome and calculate different outcome measures to describe efficacy of PrEP

	PrEP	No PrEP	Total
HIV-positive	15	40	55
HIV-negative	195	160	355
Total	210	200	410

Outcome measures

Common outcome measures for binary data:

- Risk difference
- Risk ratio
- Odds ratio

Risk difference

- Also known as the absolute risk difference
- Risk of the event occurring = the percentage of patients who experienced an event
- Risk difference is the difference between these percentages in the two groups

Risk difference

	PrEP	No PrEP	Total
HIV-positive	15	40	55
HIV-negative	195	160	355
Total	210	200	410

Risk in treatment group = 15/210 = **7.1%**
Risk in control group = 40/200 = **20%**
Risk difference = 7.1 – 20 = -**12.9%**

Risk ratio

- Instead of taking the difference of the risk in each group, we can look at their ratio:

$$\textbf{RR} = \frac{\text{risk of event in treatment group}}{\text{risk of event in control group}}$$

- Interpretation:

RR < 1: risk of event is **<u>less</u>** in treatment group than the control group

RR = 1: risk of event is the **<u>same</u>** in the treatment and control groups

RR > 1 : risk of event is **<u>greater</u>** in treatment group than the control group

Risk ratio

	PrEP	No PrEP	Total
HIV-positive	15	40	55
HIV-negative	195	160	355
Total	210	200	410

Risk in treatment group = 15/210 = **7.1%**
Risk in control group = 40/200 = **20%**
Risk ratio = 7.1/20 = 0.36

Interpreting risk

- Risk difference of -12.9%
  - There were 12.9 more HIV infections for every 100 individuals in the no PrEP group
- Risk ratio of 0.36
  - There was a 64% reduced risk of testing HIV-positive in the PrEP group (1-0.36 =0.64)
- Risk difference of 7.1-6% = 1.1%

  There were 1.1 **more** participants testing HIV-positive in the PrEP group for every 100 people
- Risk ratio of 1.18 (7.1/6)

  There was a 18% **increased** risk HIV in the PrEP group (i.e 1.18-1= 0.18*)*

Odds

- What are the odds of an event occurring?

$$\textbf{Odds} = \frac{\text{Probability of event}}{\text{Probability of no event}} = \frac{\text{Number with event}}{\text{Number with no event}}$$

A simple example – there is an 80% probability that it will rain today

Odds of raining = 0.8/0.2 = 4 (4 to 1)

Odds of not raining = 0.2/0.8 = 0.25 (1 to 4)

Odds ratio

- Instead of taking the difference of the risk in each group, we can look at their ratio:

$$\textbf{OR} = \frac{\text{odds of event in treatment group}}{\text{odds of event in control group}}$$

- Interpretation:

OR < 1: odds of event is **less** in treatment group then the control group

OR = 1: odds of event is the **same** in the treatment and control groups

OR > 1 : odds of event is **greater** in treatment group than the control group

- Hypothesis:
  - H_0: Equal risk of an event occurring

    odds ratio=1

  - H_1: Unequal risk of an event occurring

    odds ratio≠1

Odds ratios

	PrEP	No PrEP	Total
HIV-positive	**15**	**40**	55
HIV-negative	**195**	**160**	355
Total	210	200	410

Odds (HIV)	15/195	40/160	55/355
	0.08	0.25	0.15

Odds Ratio 0.08/0.25 = 0.31

Odds ratios

Odds Ratio 0.08/0.25 = 0.31

95% CI (0.15-0.60)

This does not include 1 so there is evidence to reject H_0

*There is a 69% **decrease in the odds** of acquiring HIV in the PrEP group compared to no PrEP group*

Summary

Risk difference (absolute risk):

Difference in proportions of events between the two groups

Risk ratio (relative risk):

Ratio of proportions of events between the groups

Odds ratio:

Ratio of odds (number with event/number without event) between the groups

When looking at amount of evidence to reject H_0:

Ratios

*Does 95% CI include the value **1**?*

Risk difference

*Does 95% CI include the value **0**?*

Quiz

1. Which of the following is true about prevalence and incidence?

a) Incidence refers to new cases of a disease, while prevalence refers to existing cases of a disease
b) They can both be used to measure associations between exposure and disease
c) They are both useful for establishing the determinants of disease in a population
d) All of the above

2. In a cohort study examining the association between smoking and lung cancer, suppose the risk ratio =1.5. How would you interpret this relative risk in words?

a) There were 50 more cases of lung cancer in the smokers.
b) There was a 50% increased risk of lung cancer in smokers compared to non-smokers.
c) There is no difference in risk of lung cancer between smokers and non-smokers
d) 50% of the lung cancers in smokers were due to smoking.

Answers available at end of slide set

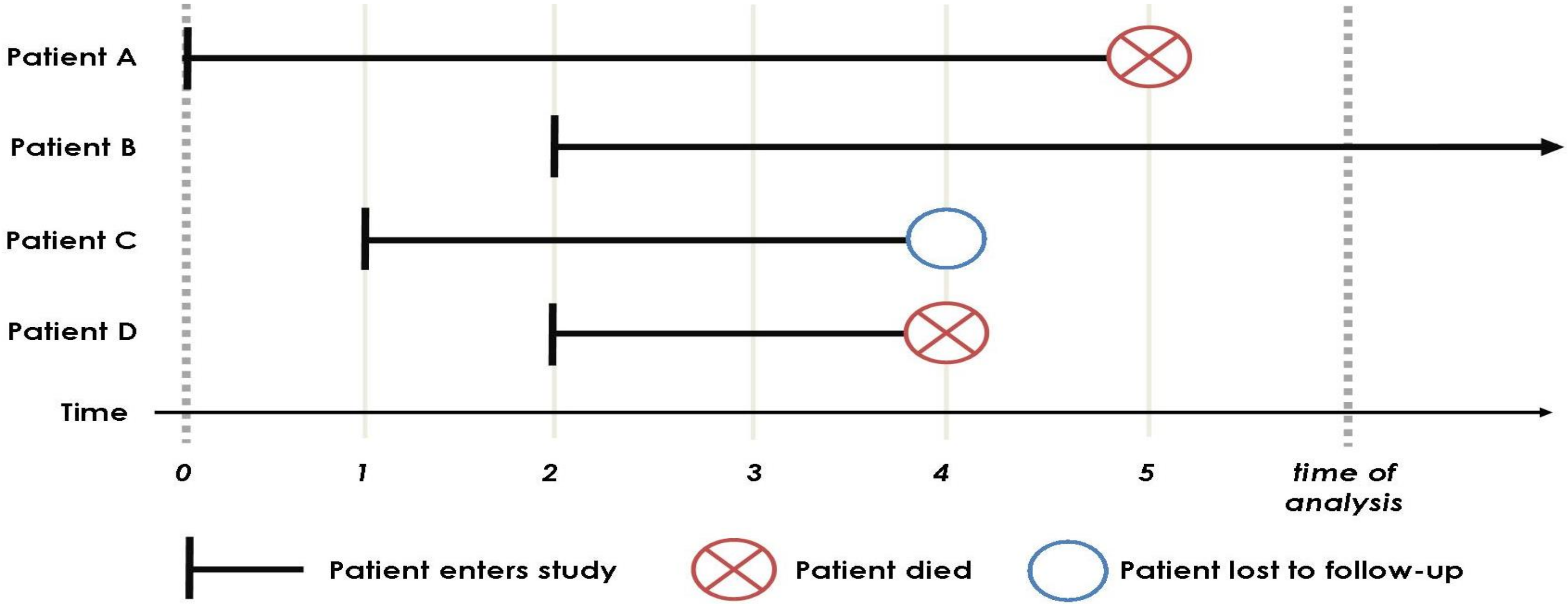
SURVIVAL ANALYSIS

DEF

Defining survival data

- In many studies, the outcome of interest is the **time to a particular event**
  - **For example:** we wish to assess the **time** from when a patient enters a clinical trial to the time a patient dies.
  - In this example we have two time-points of interest:
    - Start time: time patient entered the study
    - End time: time that patient died
  - The time between these 2 events of interest is called **survival time**

Defining survival data



Defining survival data

- Survival analysis is very frequently used in studies

- The event of interest can be:
  - negative – e.g. death, progression of disease
  - positive – e.g. discharge from the hospital
  - neutral – e.g. cessation of breast feeding

- In each case it is called survival analysis

  (sometimes time-to-event)

Example: STREAM trial

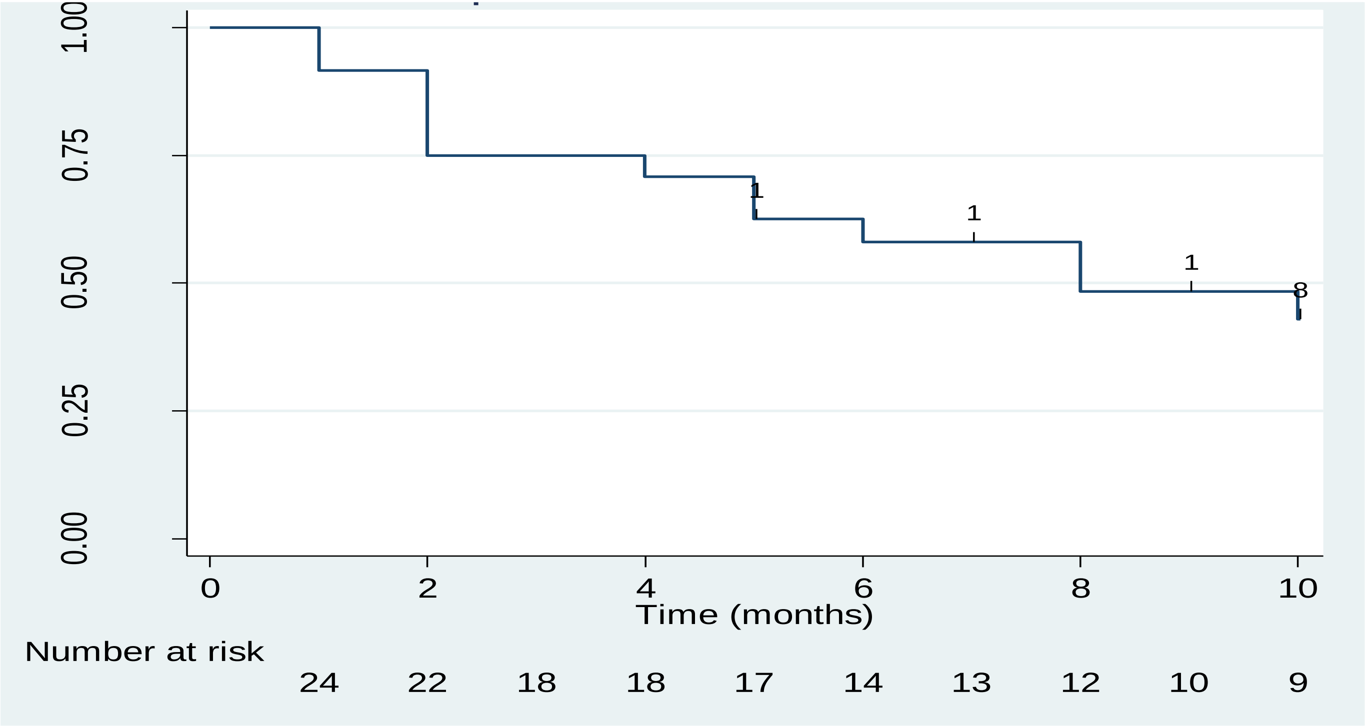
- Phase III, randomised, 2-arm, parallel-group, controlled trial
- Comparing 2 treatments for MDR-TB
- Patients treated and followed up to 132 weeks from randomisation

- Event of interest: sputum conversion (No TB detected)
- Patient's status: achieve/not achieved sputum conversion
- Starting point: date of randomisation
- Ending point: date when patient is said to have sputum converted
- Survival time: time from randomisation until sputum conversion

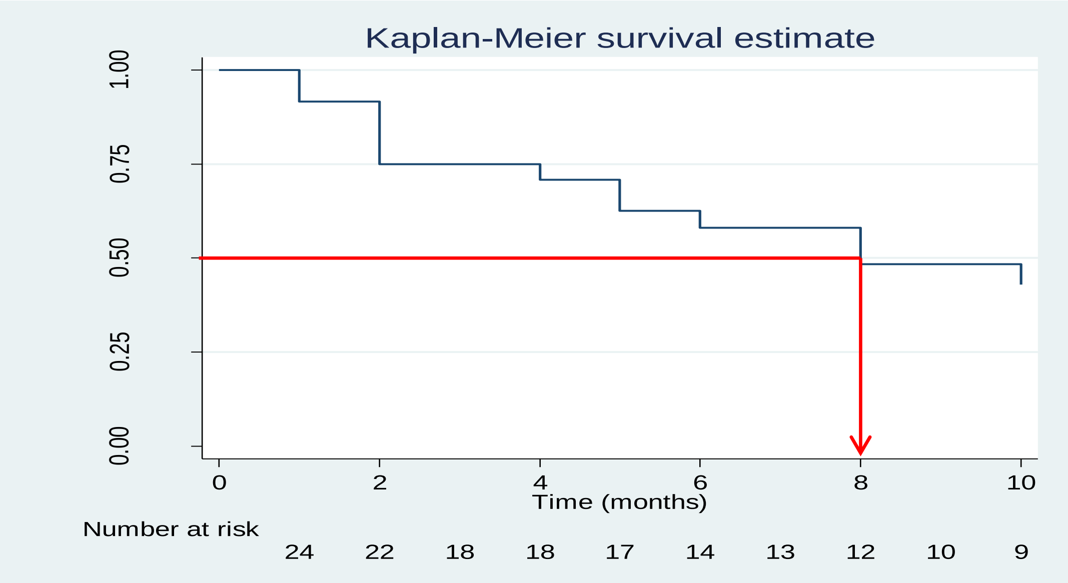
Survival function

- We can estimate the probability that a patient will survive to a certain time-point using the ‘survival function’
- The survival function estimates the probability that a patient will survive (be event-free) a certain time after some start point
- We can use the Kaplan-Meier method to estimate the survival function

Kaplan-Meier plot



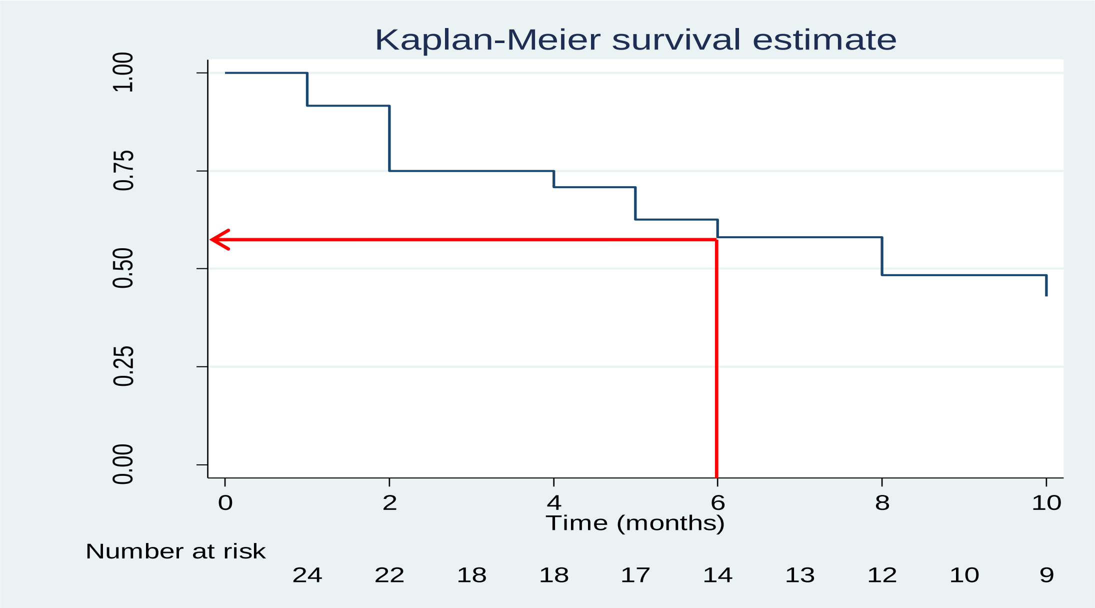
Median survival



Median survival time = **8 months**

- Value for which 50% of patients have longer survival times and 50% of patients have shorter survival times
- It is the time at which beyond 50% of patients are expected to survive (be without event)
- Can be read from the Kaplan-Meier plot
- Median survival may not be observed

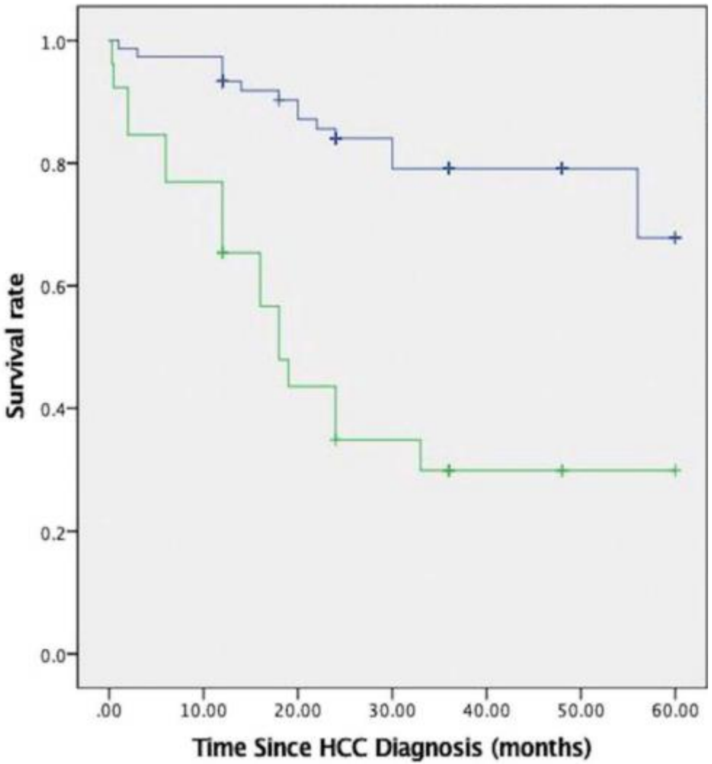
Survival rates – from graph



- Wish to estimate survival at certain time points e.g. 2 months, 4 months, 10 months
- Percentage survival at a certain time
- Read from KM plot or table

At 6 months, 58% of patients survived

Comparing groups



- Compare survival curves of two groups using **<u>log rank test</u>**
  - H_0: no difference between two groups
- Log-rank test p-value: $p<0.001$
- "There is a strong evidence of a difference in survival"

Zhao et al, JAIDS, 2021

Log-rank test limitations

- The log rank test alone gives no information about the size or direction of a difference in survival between groups, just whether there is a difference

- We also might want to know:
    - Which group sees improved survival
    - How much better / worse is the survival for a certain group

- For this we use the **'hazard ratio'**

Estimation of treatment effect

- Definition of Hazard:
  - ***the probability that a subject, having been event-free up to a certain time t, will have the event of interest within the next infinitesimal space of time***

- Hazard ratio (HR) used to estimate the difference between two survival curves

- HR is used as a measure of relative survival experience between two groups (usually experimental group vs. control)

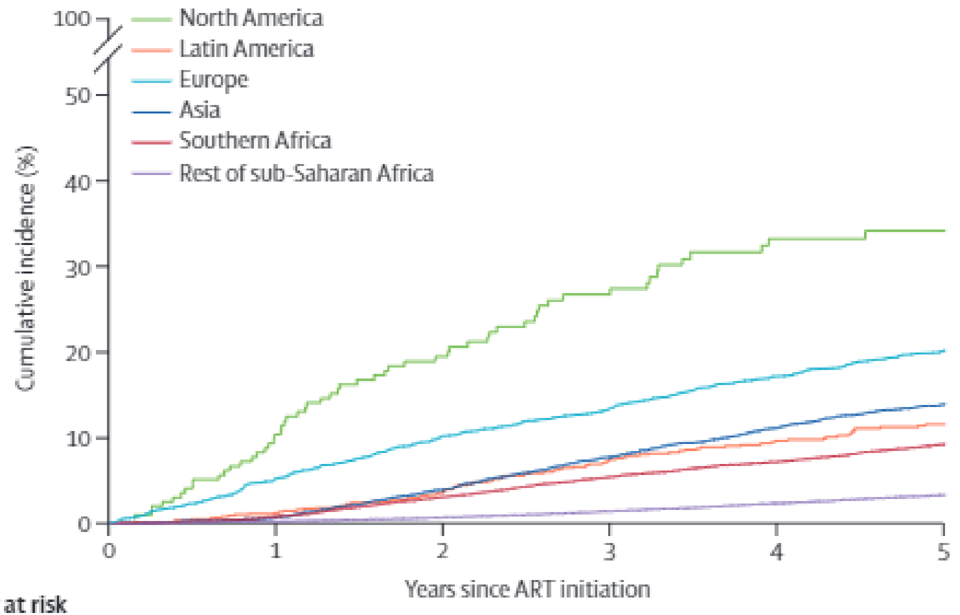
Estimation of treatment effect

- If we estimate the hazard in group 1 to be group H_1, and the hazard in group 2 to be H_2, then the ratio of the two is called the hazard ratio
  - $HR = \frac{H_1}{H_2}$
- Hence, if:
  - $HR = 1$, the risk of death is equal in both groups
  - $HR < 1$, the risk of death is less in group 1
  - $HR > 1$, the risk of death is greater in group 2

- Estimates of the HR have confidence intervals and p-values associated with them
- A confidence interval excluding 1, and p-value<0.05 implies a significant difference between the groups at the 5% level

EST

Comparing groups



Number at risk

	0	1	2	3	4	5
North America	192	167	142	108	84	68
Latin America	926	802	699	574	491	410
Europe	2142	1873	1571	1345	1086	855
Asia	6107	4950	3988	3174	2474	1914
Southern Africa	17857	13347	10171	7604	5740	4222
Rest of sub-Saharan Africa	66127	44225	32659	23535	16160	10707

Time to switch to second-line ART

Geographical region	p<0·0001
USA	4·16 (3·20–5·42)
Europe	2·30 (2·07–2·56)
Latin America	1·23 (1·03–1·49)
Asia	1·27 (1·15–1·40)
Southern Africa	1
Rest of sub-Saharan Africa	0·35 (0·33–0·38)

CIPHER Global Cohort Collaboration, Lancet HIV, 2019

Interpretation of a HR

EST

- To calculate the percentage decrease in hazard associated with being in group 1 compared to being in group 2 we can use the formula:

$$(\mathbf{1} - \boldsymbol{HR}) \times \mathbf{100}$$

- If the HR was **0.75**, then (1 – 0.75)*100 = 25%, and the interpretation would be that *‘being in group 1 was associated with a 25% decrease in hazard’*

- If the HR was **1.35**, then (1 – 1.35)*100 = -35%, and the interpretation would be that *‘being in group 1 associated with a 35% increase in hazard’*

Interpretation of a HR

For our example in the CIPHER study:

$$(1 - HR) \times 100 =$$
$$(1 - 0.35) \times 100 =$$
$$0.65 \times 100 = \mathbf{65\%}$$

Geographical region	p<0·0001
USA	4·16 (3·20–5·42)
Europe	2·30 (2·07–2·56)
Latin America	1·23 (1·03–1·49)
Asia	1·27 (1·15–1·40)
Southern Africa	1
Rest of sub-Saharan Africa	0·35 (0·33–0·38)

"being from the 'rest of sub-Saharan Africa' region was associated with a 65% reduction in the likelihood of switching to second-line treatment, compared to the 'Southern Africa' region"

EST

Interpretation of a HR

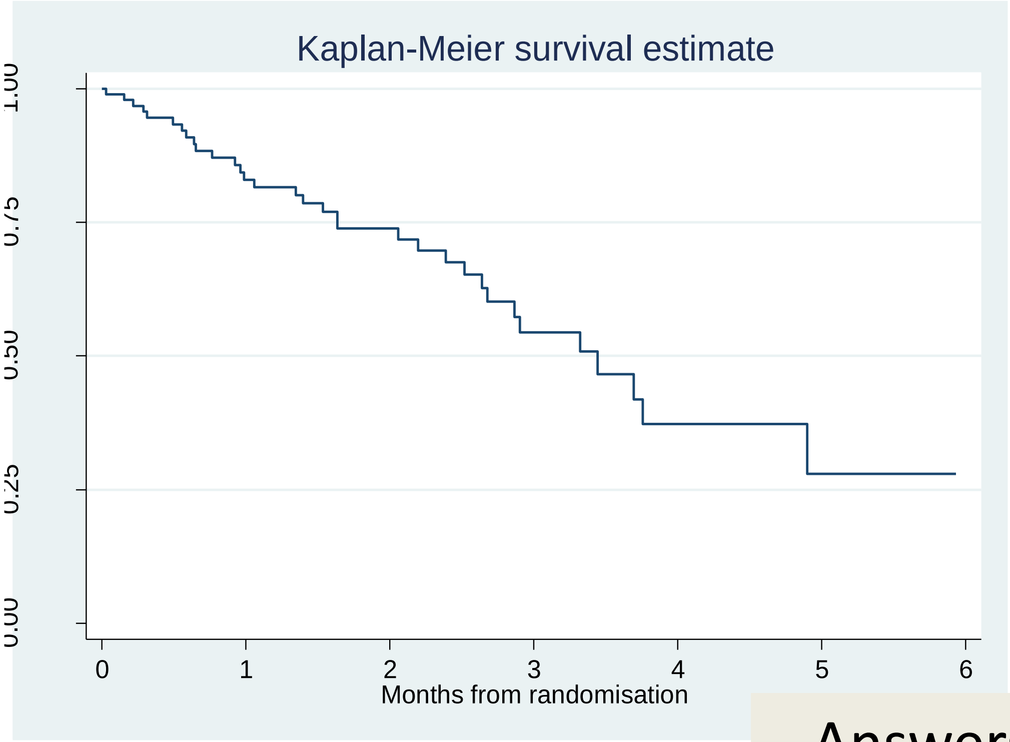
For our example in the CIPHER study:

$$(1 - HR) \times 100 =$$
$$(1 - 2.30) \times 100 =$$
$$1.30 \times 100 = \mathbf{130\%}$$

Geographical region	p<0·0001
USA	4·16 (3·20–5·42)
Europe	2·30 (2·07–2·56)
Latin America	1·23 (1·03–1·49)
Asia	1·27 (1·15–1·40)
Southern Africa	1
Rest of sub-Saharan Africa	0·35 (0·33–0·38)

"being from the 'Europe' region was associated with a 130% increase in the likelihood of switching to second-line treatment, compared to the 'Southern Africa' region"

Quiz



What is (approximately) the probability of survival to 1 month after randomisation?

a) 25%
b) 20%
c) 80%

Answers available at end of slide set

Menti question

Investigators compared how well two treatments worked preventing death for patients with cardiovascular disease. The hazard ratio they got was 1.22 (1.01, 1.47). Which of the following is the best interpretation?

a) The likelihood of death was 22% higher for patients receiving treatment A compared to B
b) The likelihood of death was 22 times higher for patients receiving treatment A compared to B
c) The likelihood of death was 122% higher for patients receiving treatment A compared to B

Answers available at end of slide set

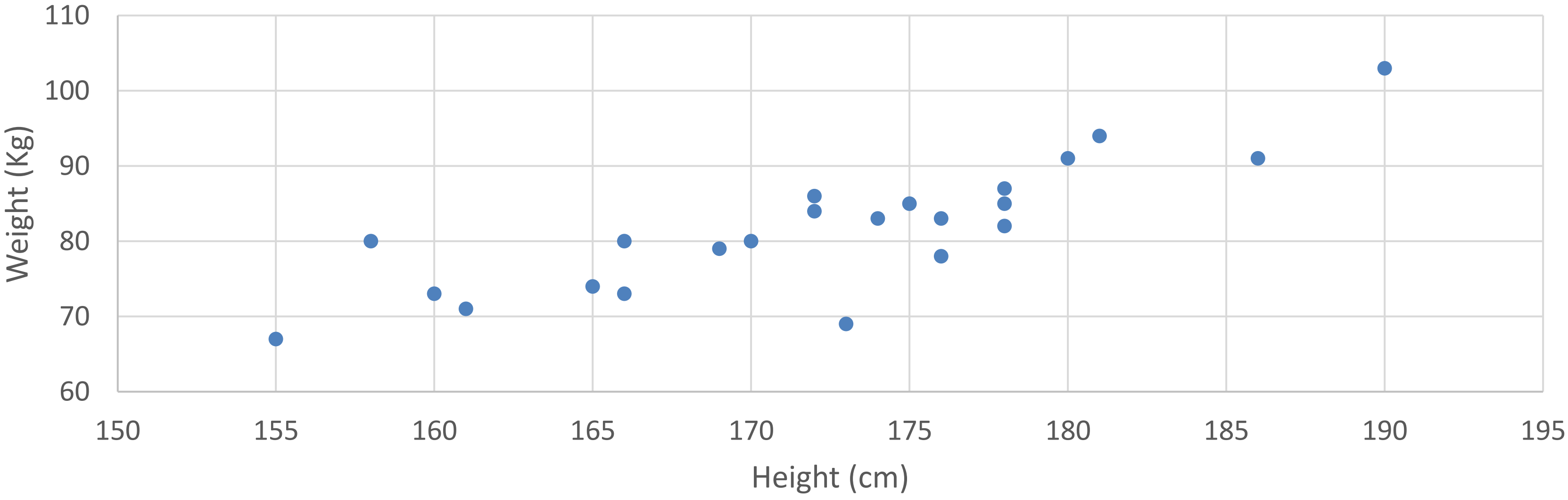
REGRESSION

What is regression?

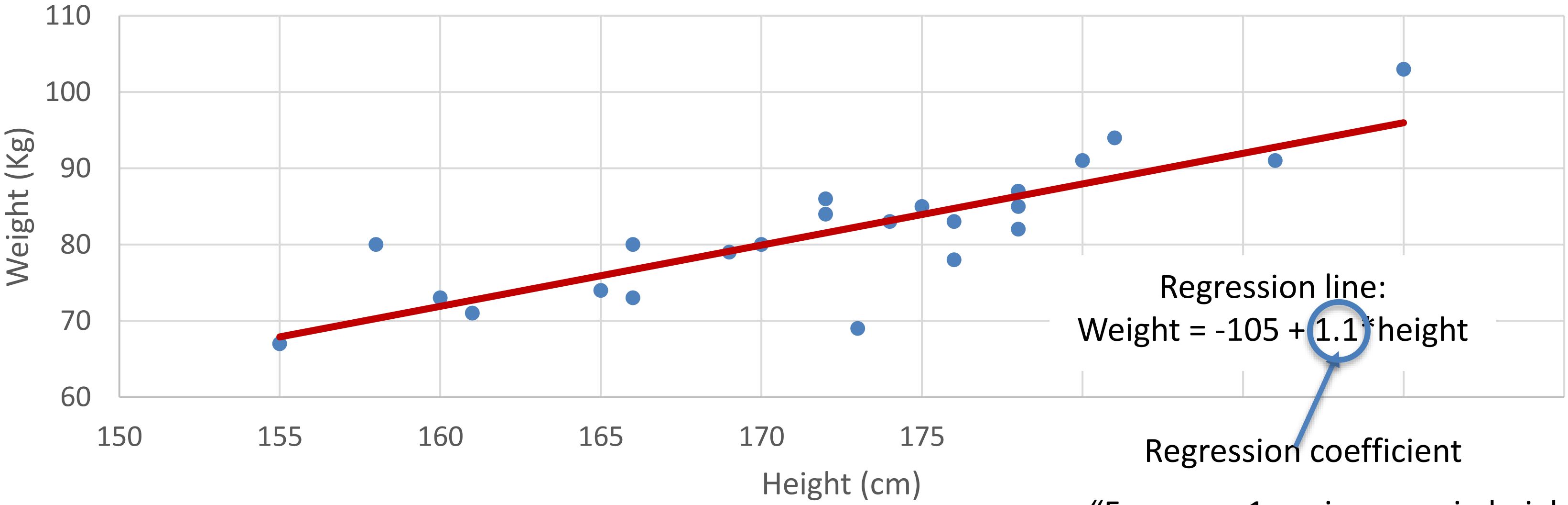
- Regression can be used to
  - Explore the impact of changes in an explanatory variable on an outcome of interest, or
  - Predict values of an outcome based on the value of one or more explanatory variables

Reminder from session 1:
- Explanatory variable = risk factor, dependent variable
- Outcome = independent variable

Example: How does weight vary by height?



Example: How does weight vary by height?



Regression line:
Weight = -105 + 1.1*height

Regression coefficient

"For every 1 cm increase in height, weight increases by an average of 1.1kg"

Examples of regression models

Type of outcome	Type of model	Interpretation of explanatory variables
Continuous	Linear regression	Coefficient (beta) = effect of a change in an explanatory variable on the mean outcome
Binary	Logistic regression	OR= effect of a change in an explanatory variable on the odds of experiencing the outcome
Survival time	Cox proportional hazards	HR= effect of a change in an explanatory variable on the hazard

Uni- or Multi-variable?

Univariable analysis = relationship between one explanatory variable and one outcome (also referred to as *bivariate* or *unadjusted* analysis)

Multivariable analysis = relationship between two or more explanatory variable and one outcome

Multivariable analysis

- Multivariable analysis can be used to ‘*adjust’* or ‘*control*’ for the effects of other variables
- Example: You are interested in whether the odds of CLWHIV being stunted at age 10 years differs between two regions
  - It is likely that age at ART initiation also differs between these regions
  - In a multivariable analysis both region and age at ART can be included in the model
    - The effect of region on odds of stunting is then said to be ‘adjusted’ or ‘controlled’ for age at ART initiation
    - The adjusted OR tells us how the odds of stunting differ between regions, if all children had initiated ART at the same age

Interpreting categorical versus continuous explanatory variables

- Continuous: coefficient represents the average change in the outcome for each one unit increase in the explanatory variable

- Categorical: coefficients represent the average difference one group compared to a reference group

Example: Multivariable regression of characteristics associated with weight (kg)

	Coefficient	95% Confidence interval	P-value
Age (years)	1.05	(1.01 to 1.09)	0.04
Height (cm)	1.10	(1.06 to 1.14)	0.01
Marital Status			
Single	Ref		<0.001
Married	2.3	(0.9 to 3.7)	
Widowed	3.4	(1.6 to 5.2)	
Divorced	3.9	(2.5 to 5.3)	

Clinical Infectious Diseases

MAJOR ARTICLE

IDSA Infectious Diseases Society of America | hivma hiv medicine association | OXFORD

Risk Factors for Coronavirus Disease 2019 (COVID-19) Death in a Population Cohort Study from the Western Cape Province, South Africa

Western Cape Department of Health in collaboration with the National Institute for Communicable Diseases, South Africa

Background. Risk factors for coronavirus disease 2019 (COVID-19) death in sub-Saharan Africa and the effects of human immunodeficiency virus (HIV) and tuberculosis on COVID-19 outcomes are unknown.

Methods. We conducted a population cohort study using linked data from adults attending public-sector health facilities in the Western Cape, South Africa. We used Cox proportional hazards models, adjusted for age, sex, location, and comorbidities, to examine the associations between HIV, tuberculosis, and COVID-19 death from 1 March to 9 June 2020 among (1) public-sector "ac-

"We conducted a population cohort study using linked data from adults attending public-sector health facilities in the Western Cape, South Africa.

We used Cox proportional hazards models, adjusted for age, sex, location, and comorbidities, to examine the associations between HIV, tuberculosis, and COVID-19 death from 1 March to 9 June 2020 among (1) public-sector "active patients" (≥1 visit in the 3 years before March 2020); (2) laboratory-diagnosed COVID-19 cases; and (3) hospitalized COVID-19 cases"

Table 2. Patient Characteristics by Human Immunodeficiency Virus Status

	Public-sector patients with HIV			Public-sector patients without HIV		
	No diagnosed COVID-19, n = 536 574	COVID-19 cases, not deceased, n = 3863	COVID-19 cases, deceased, n = 115	No diagnosed COVID-19, n = 2 902 050	COVID-19 cases, not deceased, n = 17 820	COVID-19 cases, deceased, n = 510
Sex						
Female	356 356 (66%)	3039 (79%)	62 (54%)	1 627 124 (56%)	11 877 (67%)	278 (55%)
Male	180 218 (34%)	824 (21%)	53 (46%)	1 274 926 (44%)	5943 (34%)	232 (45%)
Age						
20–39 years	310 551 (58%)	2187 (57%)	17 (15%)	1 603 235 (55%)	9453 (53%)	29 (6%)
40–49 years	147 344 (27%)	1136 (29%)	28 (24%)	457 632 (16%)	3379 (19%)	35 (7%)
50–59 years	59 345 (11%)	418 (11%)	40 (35%)	388 394 (13%)	2809 (16%)	122 (24%)
60–69 years	15 856 (3%)	98 (3%)	21 (18%)	260 226 (9%)	1325 (7%)	157 (31%)
≥70 years	3473 (1%)	24 (1%)	9 (8%)	192 562 (7%)	854 (5%)	167 (33%)
Diabetes						
None	517 609 (96%)	3491 (90%)	57 (50%)	2 659 479 (92%)	15 090 (85%)	196 (38%)
Diabetes HbA1c <7%	3493 (1%)	65 (2%)	8 (7%)	41 561 (1%)	426 (2%)	50 (10%)
Diabetes HbA1c 7–8.9%	2998 (1%)	77 (2%)	16 (14%)	44 213 (2%)	505 (3%)	78 (15%)
Diabetes HbA1c ≥9%	4562 (1%)	126 (3%)	25 (22%)	61 077 (2%)	960 (5%)	133 (26%)
Diabetes, no HbA1c measurement	7912 (1%)	104 (3%)	9 (8%)	95 720 (3%)	839 (5%)	53 (10%)
Other noncommunicable diseases						
Hypertension	62 676 (12%)	692 (18%)	48 (42%)	501 232 (18%)	4218 (24%)	314 (62%)
Chronic kidney disease	6348 (1%)	82 (2%)	21 (18%)	55 319 (2%)	412 (2%)	90 (18%)
Chronic pulmonary disease/asthma	23 501 (4%)	218 (6%)	10 (9%)	169 086 (6%)	1359 (8%)	74 (15%)
Tuberculosis						
Previous tuberculosis	129 259 (24%)	864 (22%)	42 (37%)	157 630 (5%)	834 (5%)	45 (9%)
Current tuberculosis	24 357 (5%)	172 (4%)	16 (14%)	29 895 (1%)	145 (1%)	10 (2%)

Table 3. Associations with Coronavirus Disease 2019 Death Among All Public-sector Patients ≥20 Years Old With a Public-sector Health Visit in the Previous 3 Years

	Adjusted for location only		
	HR	95% CI	*P* value
Sex			
Female	Ref	...	
Male	1.21	1.03–1.41	.02
Age			
20–39 years	Ref	...	
40–49 years	4.46	3.05–6.52	<.001
50–59 years	16.23	11.70–22.52	<.001
60–69 years	28.82	20.83–39.87	<.001
≥70 years	41.37	29.87–57.29	<.001
Diabetes			
None	Ref	...	
Diabetes HbA1c <7%	16.59	12.47–22.09	<.001
Diabetes HbA1c 7–8.9%	25.32	19.98–32.10	<.001
Diabetes HbA1c ≥9%	29.57	24.23–36.10	<.001
Diabetes, no HbA1c measurement	7.29	5.52–9.62	<.001
Other noncommunicable diseases			
Hypertension	6.72	5.73–7.88	<.001
Chronic kidney disease	11.43	9.30–14.05	<.001
Chronic pulmonary disease / asthma	2.49	1.98–3.13	<.001
Tuberculosis			
Never tuberculosis	Ref	...	
Previous tuberculosis	1.79	1.42–2.24	<.001
Current tuberculosis	2.79	1.88–4.13	<.001
HIV			
Negative	Ref	...	
Positive	1.07	.88–1.32	.494

There was no evidence of a difference in hazard of death from COVID-19 in patients with HIV and those without (after controlling for location only)

Table 3. Associations with Coronavirus Disease 2019 Death Among All Public-sector Patients ≥20 Years Old With a Public-sector Health Visit in the Previous 3 Years

	Adjusted for location only			Adjusted for age and sex		
	HR	95% CI	P value	Adjusted HR	95% CI	P value
Sex						
Female	Ref	...		Ref	...	
Male	1.21	1.03–1.41	.02	1.26	1.07–1.47	.005
Age						
20–39 years	Ref	...		Ref	...	
40–49 years	4.46	3.05–6.52	<.001	4.42	3.02–6.46	<.001
50–59 years	16.23	11.70–22.52	<.001	16.13	11.62–22.39	<.001
60–69 years	28.82	20.83–39.87	<.001	28.81	20.82–39.86	<.001
≥70 years	41.37	29.87–57.29	<.001	41.85	30.21–57.96	<.001
Diabetes						
None	Ref	...		Ref	...	
Diabetes HbA1c <7%	16.59	12.47–22.09	<.001	6.07	4.52–8.16	<.001
Diabetes HbA1c 7–8.9%	25.32	19.98–32.10	<.001	9.26	7.23–11.85	<.001
Diabetes HbA1c ≥9%	29.57	24.23–36.10	<.001	12.90	10.47–15.88	<.001
Diabetes, no HbA1c measurement	7.29	5.52–9.62	<.001	3.02	2.27–4.02	<.001
Other noncommunicable diseases						
Hypertension	6.72	5.73–7.88	<.001	2.20	1.85–2.62	<.001
Chronic kidney disease	11.43	9.30–14.05	<.001	3.21	2.57–4.01	<.001
Chronic pulmonary disease / asthma	2.49	1.98–3.13	<.001	1.08	.85–1.36	.538
Tuberculosis						
Never tuberculosis	Ref	...		Ref	...	
Previous tuberculosis	1.79	1.42–2.24	<.001	1.81	1.44–2.28	<.001
Current tuberculosis	2.79	1.88–4.13	<.001	3.29	2.21–4.88	<.001
HIV						
Negative	Ref	...		Ref	...	
Positive	1.07	.88–1.32	.494	1.97	1.59–2.45	<.001

The hazard of death from COVID-19 was 97% higher in patients with HIV than those without after adjusting for age and sex

Table 3. Associations with Coronavirus Disease 2019 Death Among All Public-sector Patients ≥20 Years Old With a Public-sector Health Visit in the Previous 3 Years

	Adjusted for location only			Adjusted for age and sex			Adjusted for all variables listed		
	HR	95% CI	*P* value	Adjusted HR	95% CI	*P* value	Adjusted HR	95% CI	*P* value
Sex									
Female	Ref	...		Ref	...		Ref	...	
Male	1.21	1.03–1.41	.02	1.26	1.07–1.47	.005	1.45	1.23–1.70	<.001
Age									
20–39 years	Ref	...		Ref	...		Ref	...	
40–49 years	4.46	3.05–6.52	<.001	4.42	3.02–6.46	<.001	2.83	1.92–4.15	<.001
50–59 years	16.23	11.70–22.52	<.001	16.13	11.62–22.39	<.001	7.78	5.51–10.98	<.001
60–69 years	28.82	20.83–39.87	<.001	28.81	20.82–39.86	<.001	11.54	8.11–16.42	<.001
≥70 years	41.37	29.87–57.29	<.001	41.85	30.21–57.96	<.001	16.79	11.69–24.11	<.001
Diabetes									
None	Ref	...		Ref	...		Ref	...	
Diabetes HbA1c <7%	16.59	12.47–22.09	<.001	6.07	4.52–8.16	<.001	5.37	3.96–7.27	<.001
Diabetes HbA1c 7–8.9%	25.32	19.98–32.10	<.001	9.26	7.23–11.85	<.001	8.53	6.60–11.02	<.001
Diabetes HbA1c ≥9%	29.57	24.23–36.10	<.001	12.90	10.47–15.88	<.001	12.07	9.70–15.02	<.001
Diabetes, no HbA1c measurement	7.29	5.52–9.62	<.001	3.02	2.27–4.02	<.001	2.91	2.18–3.89	<.001
Other noncommunicable diseases									
Hypertension	6.72	5.73–7.88	<.001	2.20	1.85–2.62	<.001	1.31	1.09–1.57	.004
Chronic kidney disease	11.43	9.30–14.05	<.001	3.21	2.57–4.01	<.001	1.86	1.49–2.33	<.001
Chronic pulmonary disease / asthma	2.49	1.98–3.13	<.001	1.08	.85–1.36	.538	.93	.73–1.17	.514
Tuberculosis									
Never tuberculosis	Ref	...		Ref	...		Ref	...	
Previous tuberculosis	1.79	1.42–2.24	<.001	1.81	1.44–2.28	<.001	1.51	1.18–1.93	.001
Current tuberculosis	2.79	1.88–4.13	<.001	3.29	2.21–4.88	<.001	2.70	1.81–4.04	<.001
HIV									
Negative	Ref	...		Ref	...		Ref	...	
Positive	1.07	.88–1.32	.494	1.97	1.59–2.45	<.001	2.14	1.70–2.70	<.001

Things to consider when interpreting an analysis

- Are all potential confounders included?

- Other sources of bias – e.g.
  - Missing data
  - Loss to follow up in a longitudinal study
  - Reporting or recall bias
  - Sampling bias - are participant representative of the population?

UCL

Quiz

Outcome = underweight at age 12 months

	Adjusted*		
	OR	95% CI	P-value
Age of mother (years)	0.96	0.94 to 0.98	0.032
HIV exposure HIV exposed – uninfected HIV unexposed HIV exposed - infected	 1 0.51 2.51	 0.25 to 0.68 1.50 to 3.02	 0.043
*also adjusted for child sex, birthweight, socio-economic status, region			

What is the association between age of the mother and odds of the infant being underweight at aged 12 months?

A. No evidence of an association

B. As mothers age increases, the odds of being underweight decreases

C. As mothers age increases, the odds of being underweight increases

Quiz

Outcome = underweight at age 12 months

	Adjusted*		
	OR	95% CI	P-value
Age of mother (years)	0.96	0.94 to 0.98	0.032
HIV exposure HIV exposed – uninfected HIV unexposed HIV exposed - infected	 1 0.51 2.51	 0.25 to 0.68 1.50 to 3.02	 0.043
*also adjusted for child sex, birthweight, socio-economic status, region			

Which group have the highest odds of being underweight at 12 months

A. HIV exposed, uninfected infants

B. HIV unexposed infants

C. HIV exposed, infected infants

Quiz questions and answers

Quiz

1. Which of the following is true about prevalence and incidence?

a) Incidence refers to new cases of a disease, while prevalence refers to existing cases of a disease

b) They can both be used to measure associations between exposure and disease

c) They are both useful for establishing the determinants of disease in a population

d) All of the above

2. In a cohort study examining the association between smoking and lung cancer, suppose the risk ratio =1.5. How would you interpret this relative risk in words?

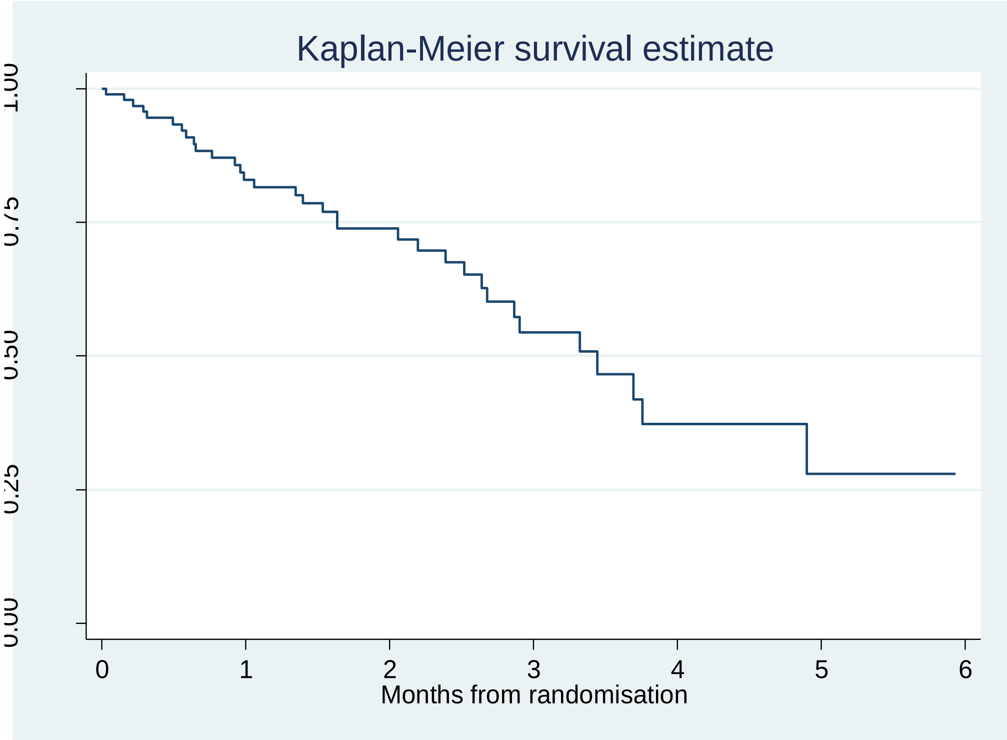
a) There were 50 more cases of lung cancer in the smokers.

b) There was a 50% increased risk of lung cancer in smokers compared to non-smokers.

c) There is no difference in risk of lung cancer between smokers and non-smokers

d) 50% of the lung cancers in smokers were due to smoking.

Quiz



What is (approximately) the probability of survival to 1 month after randomisation?

a) 25%
b) 20%
c) 80%

Quiz

Investigators compared how well two treatments worked preventing death for patients with cardiovascular disease. The hazard ratio they got was 1.22 (1.01, 1.47). Which of the following is the best interpretation?

a) The likelihood of death was 22% higher for patients receiving treatment A compared to B

b) The likelihood of death was 22 times higher for patients receiving treatment A compared to B

c) The likelihood of death was 122% higher for patients receiving treatment A compared to B

Quiz

Outcome = underweight at age 12 months

	Adjusted*		
	OR	95% CI	P-value
Age of mother (years)	0.96	0.94 to 0.98	0.032
HIV exposure HIV exposed – uninfected HIV unexposed HIV exposed - infected	 1 0.51 2.51	 0.25 to 0.68 1.50 to 3.02	 0.043
*also adjusted for child sex, birthweight, socio-economic status, region			

What is the association between age of the mother and odds of the infant being underweight at aged 12 months?

A. No evidence of an association

B. As mothers age increases, the odds of being underweight decreases

C. As mothers age increases, the odds of being underweight increases (p-value <0.05 providing evidence of a statistically significant association. OR<1 tells us as age increase, odds of being underweight decreases)

Quiz

Outcome = underweight at age 12 months

	Adjusted*		
	OR	95% CI	P-value
Age of mother (years)	0.96	0.94 to 0.98	0.032
HIV exposure HIV exposed – uninfected HIV unexposed HIV exposed - infected	 1 0.51 2.51	 0.25 to 0.68 1.50 to 3.02	 0.043
*also adjusted for child sex, birthweight, socio-economic status, region			

Which group have the highest odds of being underweight at 12 months

A. HIV exposed, uninfected infants

B. HIV unexposed infants

C. HIV exposed, infected infants HIV exposed uninfected are the reference group. Compared to the reference HIV unexposed have lower odds (as OR<1) but HIV infected have higher odds (OR>1)