EACS HIV Summer School 2017

Developing an analysis plan

Valentina Cambiano,
Institute for Global Health, University College London

Conflict of Interests

I received personal fees from Merck Sharp & Dohme Limited in 2015.

Outline of Session

- A real-life example:
  1. The idea
  2. Which data to use
  3. Are the date secure?
  4. Are the data we need available?
  5. Refining the research question
  6. Consideration of biases and limitations
- The stages of an analysis plan

1 - The idea (1)

- A clinician at a London-based HIV clinic who was involved in Spartac RCT was interested in whether treatment toxicities and antiretroviral resistance were more likely to occur in those starting cART with high CD4 counts
- Secondary end-points in Spartac and HPTN 052 (START yet to report)
- *"Would it be possible to investigate this using the routine clinic data collected through UK CHIC?"*

1 – The idea (2)

Q: *Are treatment toxicities and antiretroviral resistance more likely to occur in those starting cART with high CD4 counts?*

Population

People living with HIV initiating antiretroviral therapy for the first time

Exposure ('Intervention'/Comparator group)

I = People starting cART with a 'high' CD4 count

C= People starting cART with a 'not high' CD4 count

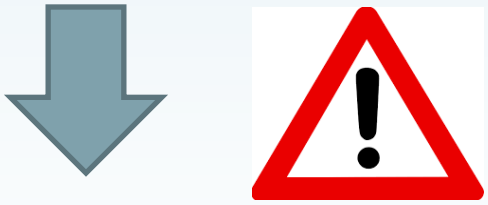
Outcome

Antiretroviral toxicities

Antiretroviral resistance

2 – Which data to use? Example

- UK CHIC is a **cohort study** which collates data on people living with HIV aged over 16 years who have accessed care in one of the collaborating centres in the UK, at any time in 1996 or thereafter.
- **Data** are those collected in **routine** HIV clinical care (*CD4 counts, HIV viral load, ART, basic demographics)*



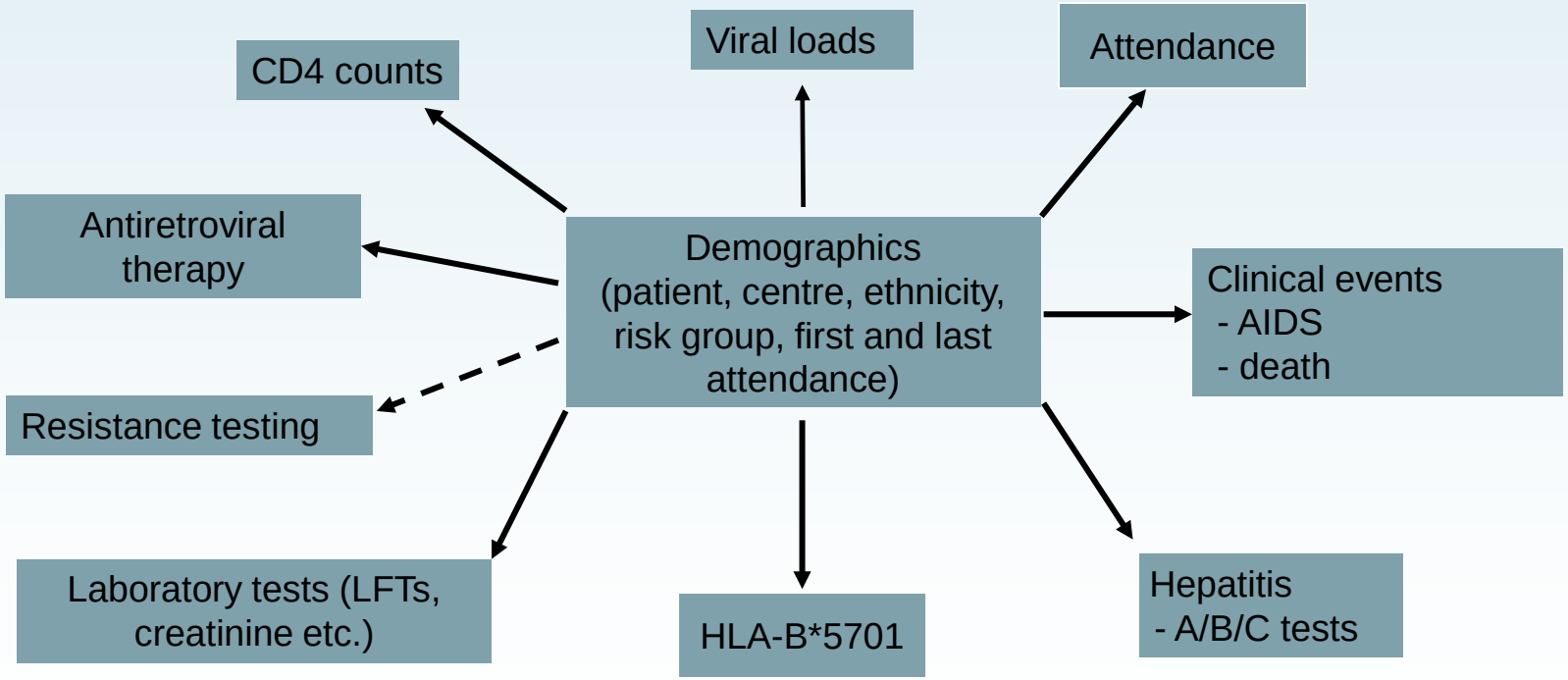
1. Will the data be secure?
2. Pragmatically can I answer this research question?
3. what are the potential limitations of these data?

3 – Will the data be secure? Example

- UK CHIC Study set up and approved for research purposes
- The data are pseudonomysed and doesn't require informed consent
- Need to check local policies and procedures for obtaining ethical approval before proceeding

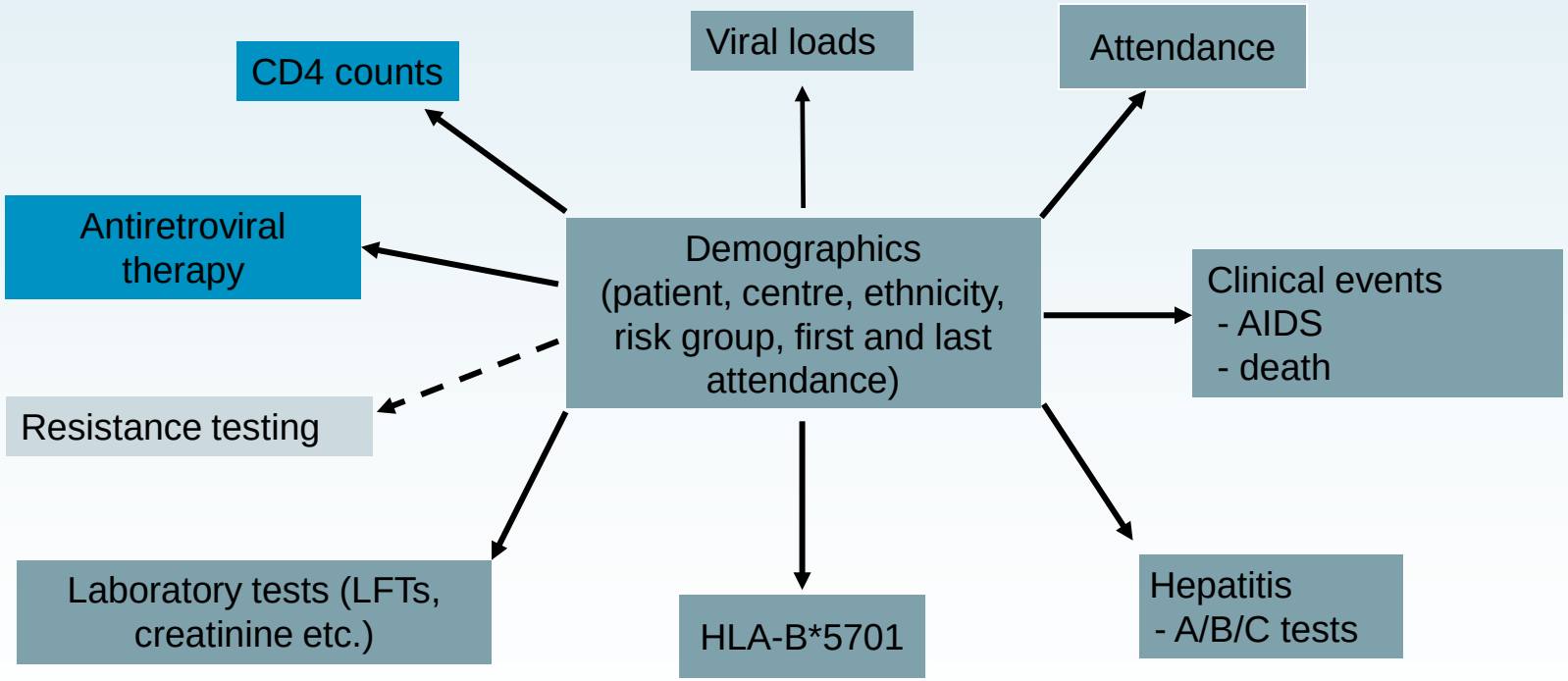
4 – Are the data we need available? (1)

Q: Are treatment toxicities and antiretroviral resistance more likely to occur in those starting cART with high CD4 counts?



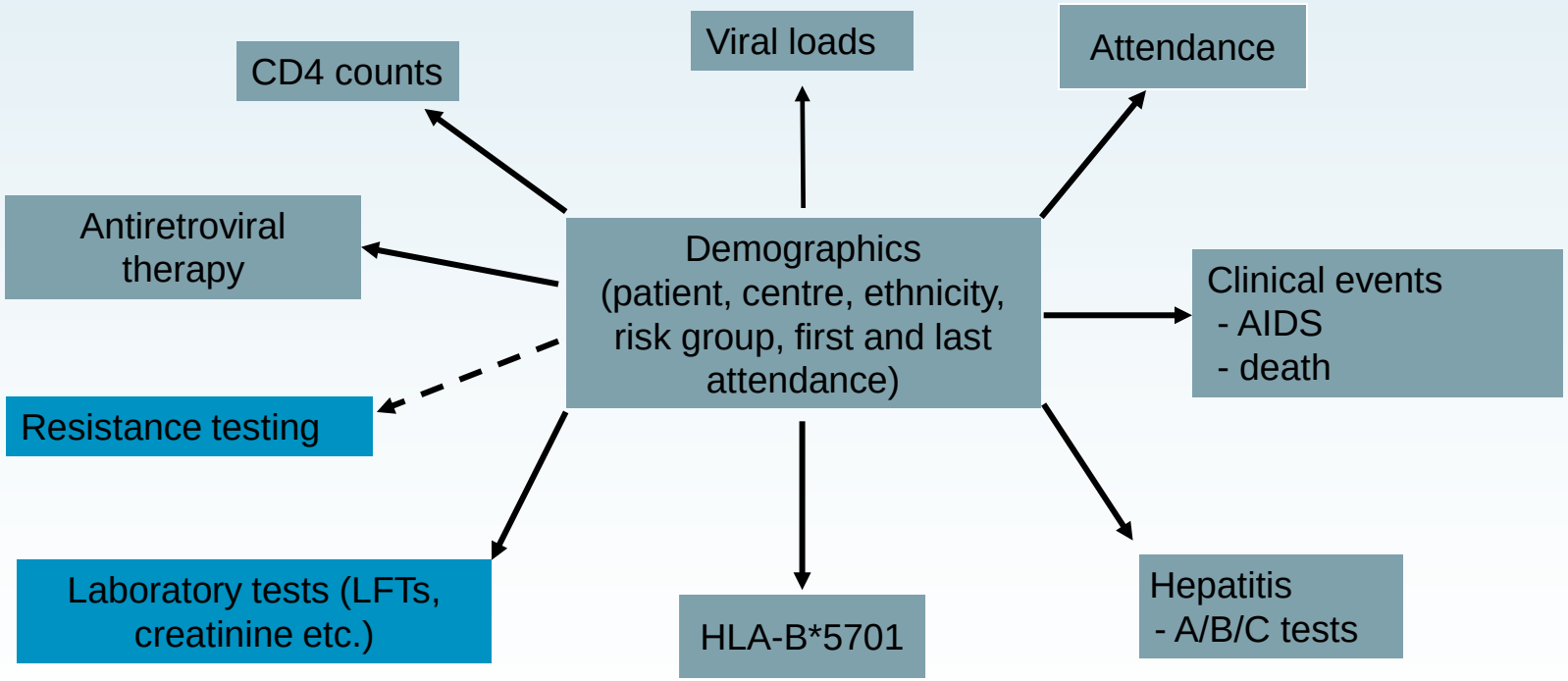
4 – Are the data we need available? (2)

Q: Are treatment toxicities and antiretroviral resistance more likely to occur in those starting cART with high CD4 counts?



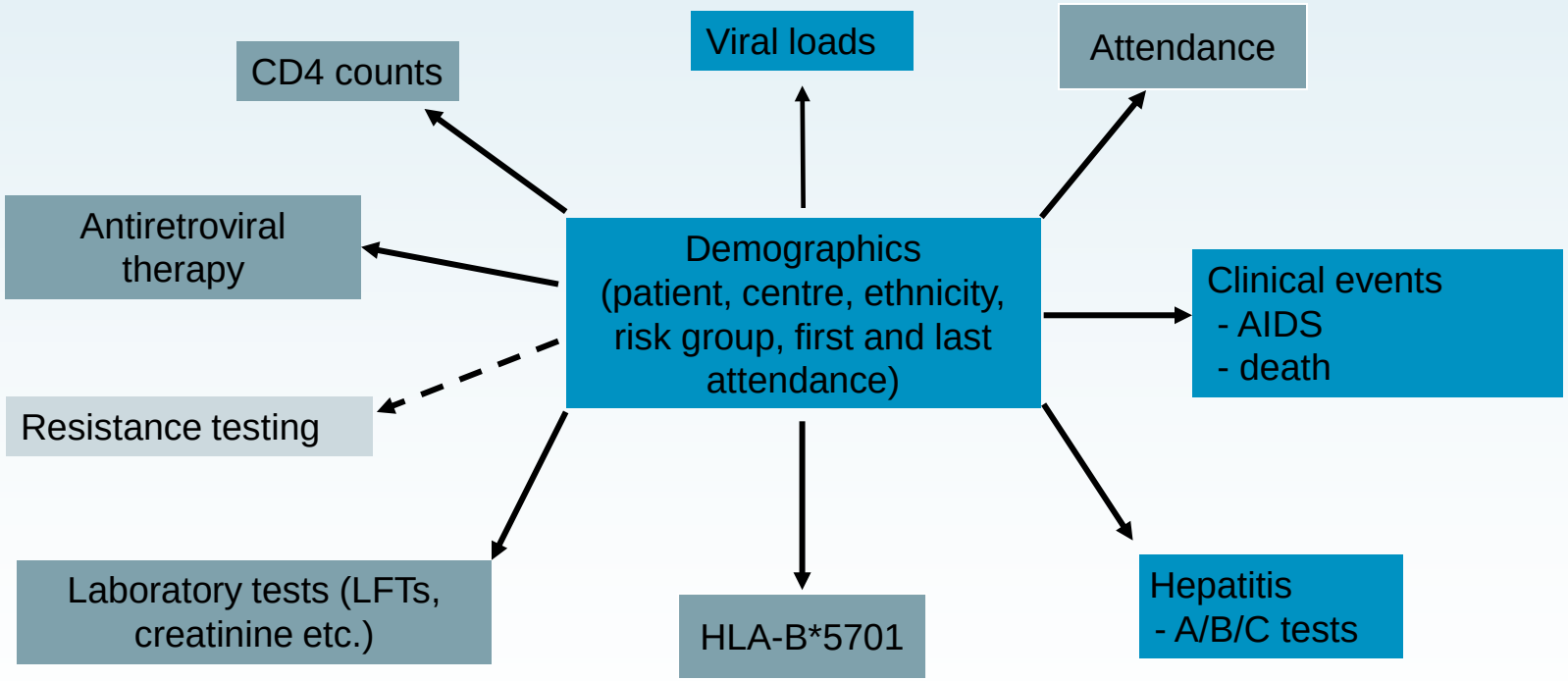
4 – Are the data we need available? (3)

Q: Are treatment toxicities and antiretroviral resistance more likely to occur in those starting cART with high CD4 counts?



4 – Are the data we need available? (4)

Q: Are treatment toxicities and antiretroviral resistance more likely to occur in those starting cART with high CD4 counts?



5 – Refining the research question (1)

Q: Are treatment toxicities and antiretroviral resistance more likely to occur in those starting cART with high CD4 counts?

Population

- HIV-positive adults initiating cART for first time
- Also need to have laboratory tests, resistance tests, CD4 counts and viral load available
- Exclusions criteria: pregnancy and cART started before 2000

5 – Refining the research question (2)

How do we measure exposure?

- What constitutes ‘high’ CD4 count?

Exposure

- People starting cART at CD4 count:
  - ≥500 cells/mm^3
  - 350-499 cells/mm^3
  - <350 cells/mm

5 – Refining the research question (3)

How do we measure outcome? Toxicities

- It could have been treatment discontinuations/ reported side-effects, etc..
- In practice, we had to choose laboratory abnormalities. Grade 3/4 adverse events[1] in any of the following laboratory markers:

Liver	Renal	Blood	Other
ALT	Creatinine	Haemoglobin	Amylase
AST		Platelet count	Cholesterol
ALP			Glucose
Albumin			

[1] Division of AIDS Table for Grading the Severity of Adult and Paediatric Adverse Events

5 – Refining the research question (4)

How do we measure outcome? Resistance

- How do we define antiretroviral resistance? (new mutations/ susceptibility scores ?)
- New PI, NNRTI or NRTI resistance mutations (compared to baseline) upon virological failure (as this is the time when a resistance test is performed)

5 – Refining the research question (5)

Q: Are treatment toxicities and antiretroviral resistance more likely to occur in those starting cART with high CD4 counts?

Do/Are ART-naïve HIV-positive adults initiating cART with CD4 counts above 350 cells/mm^3 :

Q1: experience a higher rate of laboratory-defined adverse events on cART than those initiating cART with CD4 counts of 350 cells/mm^3 and below?

Q2: more likely to develop antiretroviral resistance than those initiating cART with CD4 counts of 350 cells/mm^3 and below?

6 – What biases/limitations do we need to consider when using routinely collected data? (1)

Confounding

1. Before the guidelines were changed, who would have started cART with high CD4 counts?

2. Do you think these people would differ from those starting with low CD4 count? How?

3. Have these data been collected (so that we can adjust for these factors)?

6 – What biases do we need to consider when using routinely collected data? (2)

		CD4 count at start of ART (cells/mm³)		
		≤350	351-499	≥500
Sex, n (%)	Male	6147 (78.2)	958 (87.2)	406 (90.8)
Ethnicity, n (%)	White	4586 (58.4)	787 (71.6)	334 (74.7)
	Black African	1861 (23.7)	131 (11.9)	42 (9.4)
	Black other	404 (5.1)	52 (4.7)	17 (3.8)
	Other/unknown	1009 (12.8)	129 (11.7)	54 (12.1)
Mode of HIV acquisition, n (%)	Sex between men	4518 (57.5)	801 (72.9)	347 (77.6)
	Heterosexual	2739 (34.9)	212 (19.3)	68 (15.2)
	Other/unknown	603 (7.7)	86 (7.8)	32 (7.2)
Regimen type, n (%)	2 NRTI + PI (/r)	1893 (24.1)	311 (28.3)	186 (41.6)
	2 NRTI + NNRTI	5559 (70.7)	718 (65.3)	236 (52.8)
	≥ 3 NRTI	173 (2.2)	22 (2.0)	8 (1.8)
	Other Combination	235 (3.0)	48 (4.4)	17 (3.8)

6 – What biases do we need to consider when using routinely collected data? (3)

		CD4 count at start of ART (cells/mm³)		
		≤350	351-499	≥500
Sex, n (%)	Male	6147 (78.2)	958 (87.2)	406 (90.8)
Ethnicity, n (%)	White	4586 (58.4)	787 (71.6)	334 (74.7)
	Black African	1861 (23.7)	131 (11.9)	42 (9.4)
	Black other	404 (5.1)	52 (4.7)	17 (3.8)
	Other/unknown	1009 (12.8)	129 (11.7)	54 (12.1)
Mode of HIV acquisition, n (%)	Sex between men	4518 (57.5)	801 (72.9)	347 (77.6)
	Heterosexual	2739 (34.9)	212 (19.3)	68 (15.2)
	Other/unknown	603 (7.7)	86 (7.8)	32 (7.2)
Regimen type, n (%)	2 NRTI + PI (/r)	1893 (24.1)	311 (28.3)	186 (41.6)
	2 NRTI + NNRTI	5559 (70.7)	718 (65.3)	236 (52.8)
	≥ 3 NRTI	173 (2.2)	22 (2.0)	8 (1.8)
	Other Combination	235 (3.0)	48 (4.4)	17 (3.8)[19]

6 – What biases do we need to consider when using routinely collected data? (4)

		CD4 count at start of ART (cells/mm³)		
		≤350	351-499	≥500
Sex, n (%)	Male	6147 (78.2)	958 (87.2)	406 (90.8)
Ethnicity, n (%)	White	4586 (58.4)	787 (71.6)	334 (74.7)
	Black African	1861 (23.7)	131 (11.9)	42 (9.4)
	Black other	404 (5.1)	52 (4.7)	17 (3.8)
	Other/unknown	1009 (12.8)	129 (11.7)	54 (12.1)
Mode of HIV acquisition, n (%)	Sex between men	4518 (57.5)	801 (72.9)	347 (77.6)
	Heterosexual	2739 (34.9)	212 (19.3)	68 (15.2)
	Other/unknown	603 (7.7)	86 (7.8)	32 (7.2)
Regimen type, n (%)	2 NRTI + PI (/r)	1893 (24.1)	311 (28.3)	186 (41.6)
	2 NRTI + NNRTI	5559 (70.7)	718 (65.3)	236 (52.8)
	≥ 3 NRTI	173 (2.2)	22 (2.0)	8 (1.8)
	Other Combination	235 (3.0)	48 (4.4)	17 (3.8)[20]

6 – What biases/limitations do we need to consider when using routinely collected data? (5)

Missing data

1. If we think about the exposure, in this case CD4 at cART start, can you think of why some people may be less likely to have it recorded?

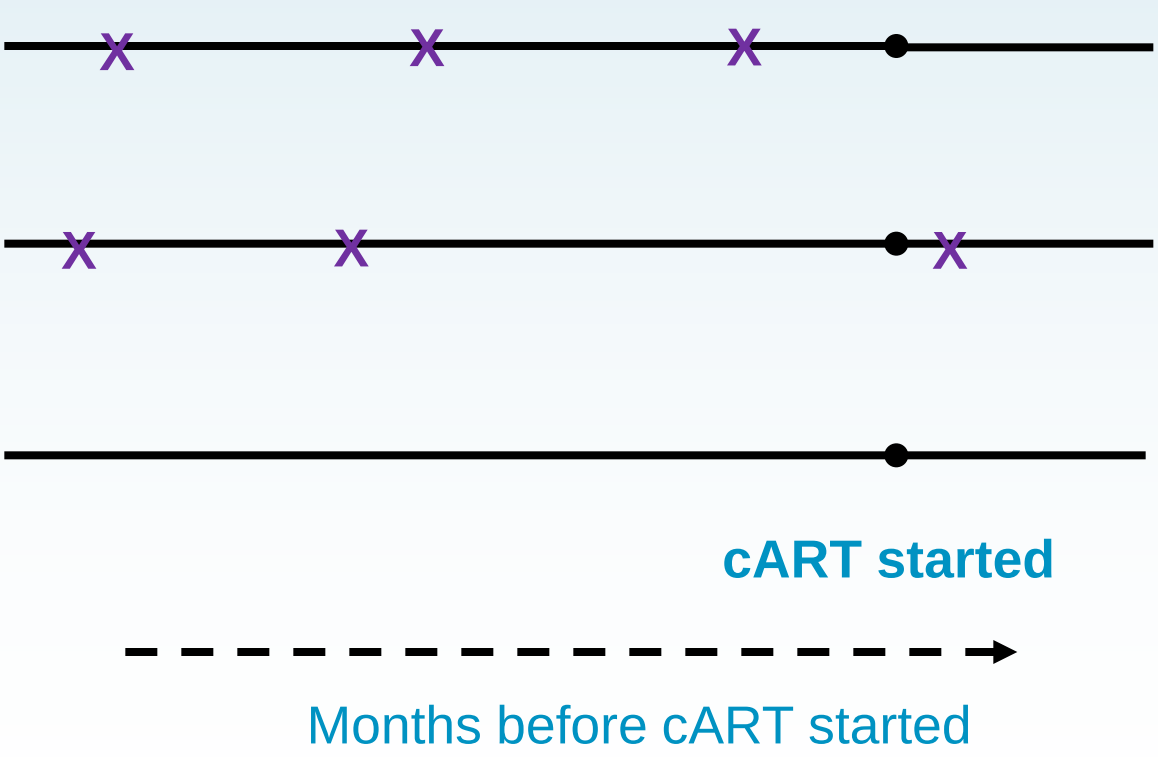
2. What about the outcome? Who is going to have missing value for the outcome (in this case lab toxicities or resistance)?

6 – What biases/limitations do we need to consider when using routinely collected data? (6)

Months before cART started

X – CD4 count measured

6 – What biases/limitations do we need to consider when using routinely collected data? (7)



What would you do with this people? How would you treat them in the analysis?

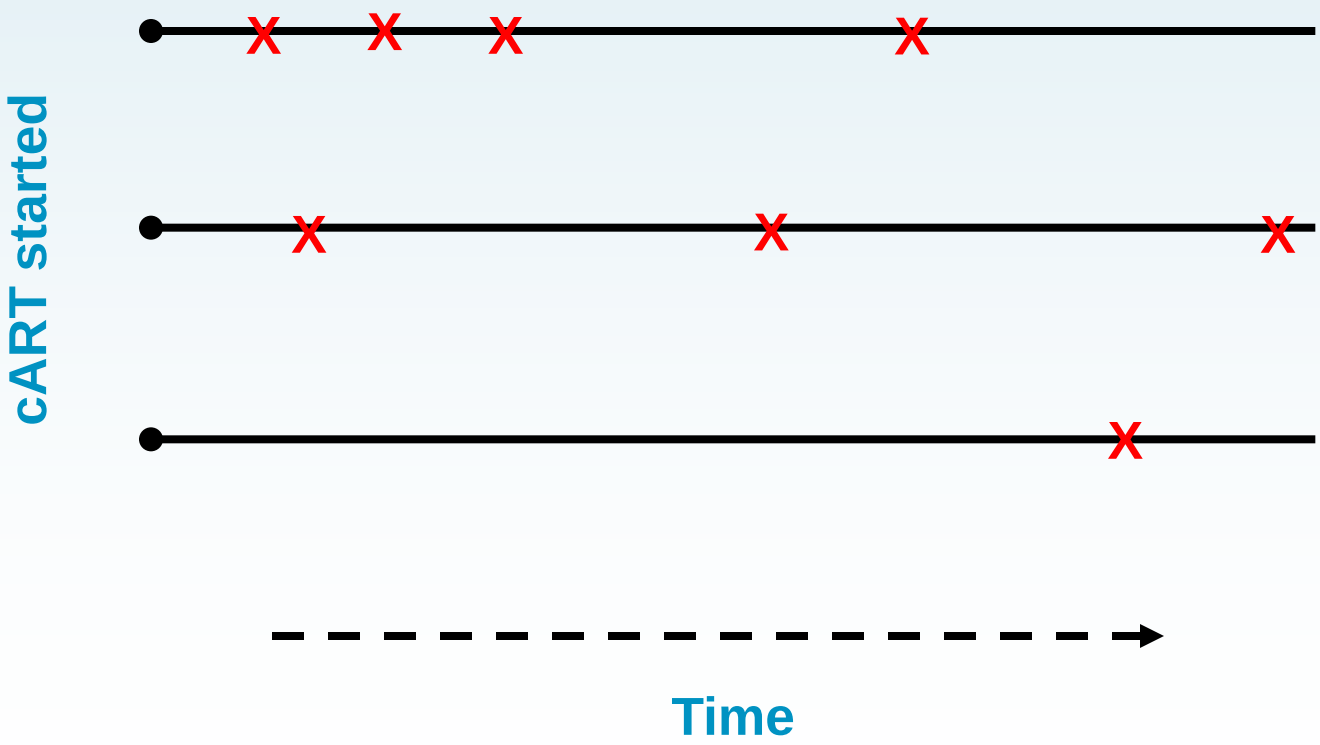
X – CD4 count measured

6 – What biases/limitations do we need to consider when using routinely collected data? (8)

- **Infrequent monitoring**
  - Laboratory tests not performed at regular intervals in all patients

1. Can you think of any reasons why some people will have more lab measurements than others? If these data are routinely collected

6 – What biases/limitations do we need to consider when using routinely collected data? (7)



Outline of Session

- A real-life example:
  1. The idea
  2. Which data to use?
  3. Are the data secure?
  4. Is the data we need available?
  5. Refining the research question
  6. Consideration of biases and limitations
- The stages of an analysis plan

Before writing the analysis plan (1)

1. **Feasibility analysis**
   - Need to have enough people starting cART at CD4 counts above 350 cells/mm^3
   - Patients need to have:
     - laboratory measures available
     - resistance test results available at cART start and virological failure
   - Will our sample size be large enough?
   - Will we observe enough outcome events?

Before writing the analysis plan (2)

2. Write the concept sheet

1-1.5 pages

- Title
- Hypothesis/background
- Aim(s)
- Inclusion/exclusion
- Proposed analysis in brief
- Variables needed
- Possible limitations

Before writing the analysis plan (3)

2. Write the concept sheet

1-1.5 pages

- Title ✓
- Hypothesis/background ✓
- Aim(s) ✓
- Inclusion/exclusion ✓
- Proposed analysis in brief ✗
- Variables needed ✓
- Possible limitations ✓

Analysis plan (1)

- Research question and choice of exposure and outcome variables will determine most appropriate statistical tests/models to use
- Should be outlined in concept sheet
- Sensitivity and sub-group analyses may be needed – should be planned a priori

But . . . .

- May have to be more flexible than in RCTs
- As project develops, analysis plan may change

Analysis plan (2)

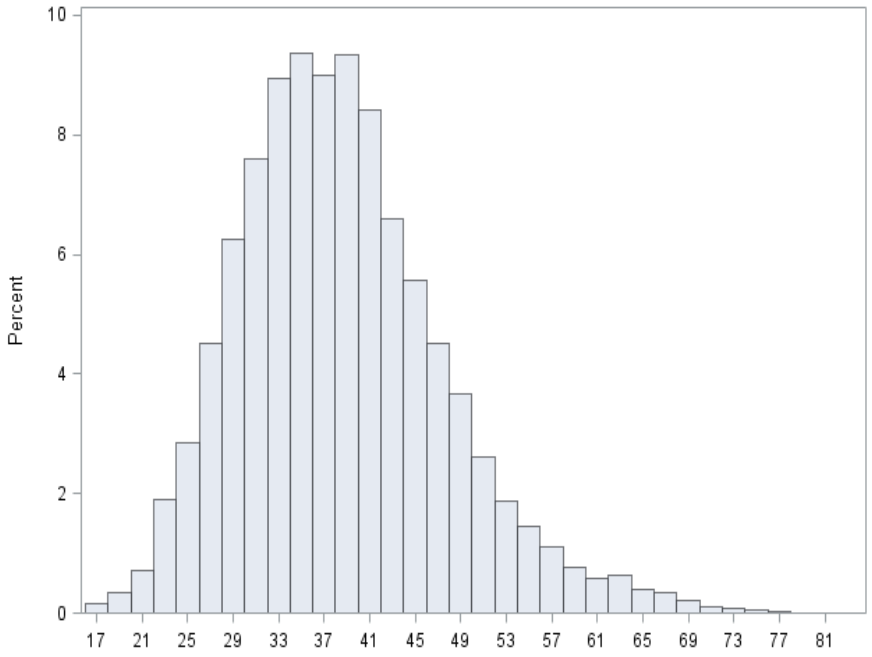
- Descriptive analyses
  - Get to know your data!
  - Identify differences in exposure groups and potential confounders
- Exploratory analyses
  - Not your main end-point
  - Provide some insight/aids interpretation of main results
- Main analysis
  - Analysis of primary end-point
  - Adjusting for confounders (regression models)
- Sensitivity and sub-group analyses
  - Are methods valid?

Descriptive analyses (1)

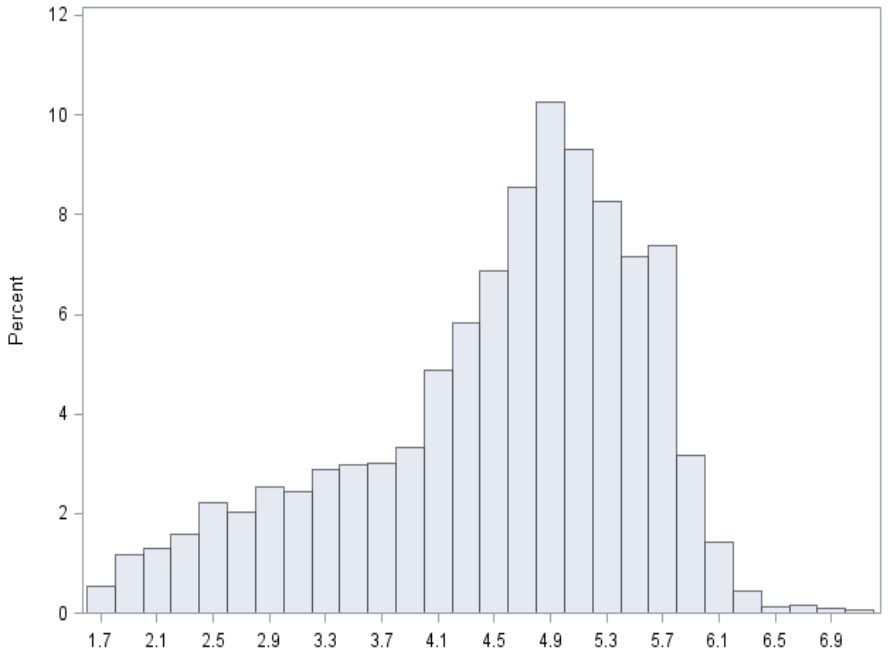
1. Data checks
2. Patient flow through study
3. Number of people with and without exposure of interest
4. Describe baseline demographic and clinical characteristics
   - According to exposure of interest
   - Appropriate univariate tests
5. Patients follow-up
6. Outcome

Descriptive analyses (2)

- 1. Data checks (errors, outliers, normal distribution, missing data etc.)



Age at cART start

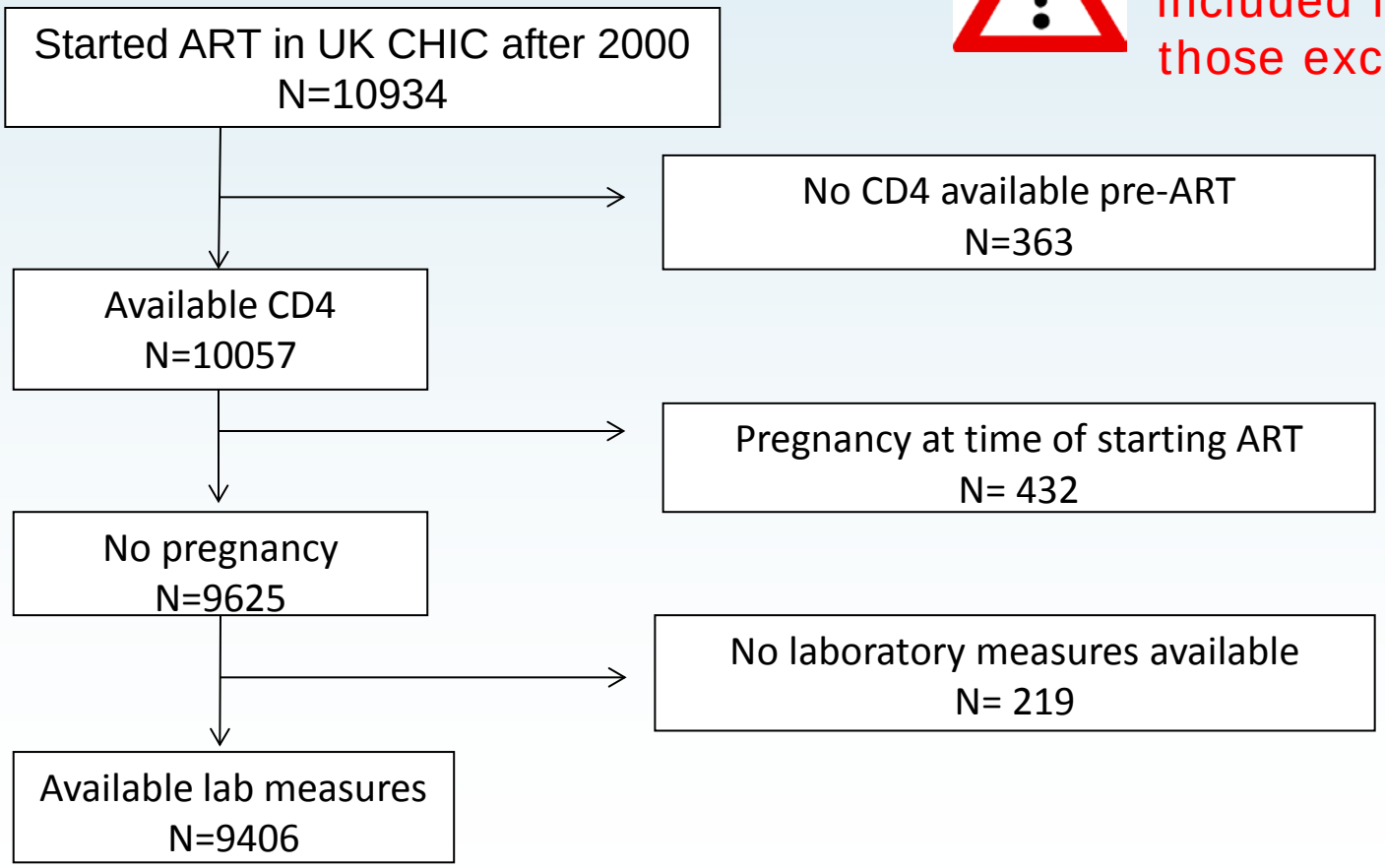


Log_{10} VL at cART start

Descriptive analyses (3)

- 2. Patient flow-chart through study

Compare those included from those excluded



Descriptive analyses (3)

- 3. Number of people with and without exposure of interest
- 4. Other characteristics of the sample
  - Demographics, HIV markers, AIDS events, cART regimen, calendar year, HBV/HCV co-infection etc.
  - Comparison by exposure of interest (CD4 count groups)
  - Univariate tests (chi-square, t-test, ANOVA etc.)

Descriptive analyses (4)

- 4. Baseline characteristics

		CD4 count at start of ART (cells/mm³)			P-value
		≤350	351-499	≥500	
Sex, n (%)	Male	6147 (78.2)	958 (87.2)	406 (90.8)	<0.001
Ethnicity, n (%)	White	4586 (58.4)	787 (71.6)	334 (74.7)	<0.001
	Black African	1861 (23.7)	131 (11.9)	42 (9.4)	
	Black other	404 (5.1)	52 (4.7)	17 (3.8)	
	Other/unknown	1009 (12.8)	129 (11.7)	54 (12.1)	
Mode of HIV acquisition, n (%)	Sex between men	4518 (57.5)	801 (72.9)	347 (77.6)	<0.001
	Heterosexual	2739 (34.9)	212 (19.3)	68 (15.2)	
	Other/unknown	603 (7.7)	86 (7.8)	32 (7.2)	
Regimen type, n (%)	2 NRTI + PI (/r)	1893 (24.1)	311 (28.3)	186 (41.6)	<0.001
	2 NRTI + NNRTI	5559 (70.7)	718 (65.3)	236 (52.8)	
	≥ 3 NRTI	173 (2.2)	22 (2.0)	8 (1.8)	
	Other Combination	235 (3.0)	48 (4.4)	17 (3.8)	

Descriptive analyses (5)

- 5. Patient follow-up
  - Time under follow-up
  - Frequency of monitoring (laboratory tests, resistance tests)
  - Comparison by exposure of interest (CD4 count groups)
  - Univariate tests

		CD4 count at start of ART (cells/mm^3)		
		≤350	351-499	≥500
Patient follow up, years	Sum	24628.0	2330.4	902.3
	Median (IQR)	2.5 (0.0, 11)	1.4 (0.0, 10.6)	1.0 (0.0, 10.6)

Descriptive analyses (6)

- 6. Outcome
  - Crude rate of grade 3/4 laboratory adverse events
  - Number of resistance mutations at viral failure
  - Comparison by exposure of interest
  - Univariate tests

Baseline CD4 count, cells/mm³	N	Virological rebound, n (%)	Resistance test, n (%)	New resistance mutation, n (%)
≤350	6514	488 (7.5)	260 (53.3)	107 (41.2)
351-499	996	46 (4.6)	20 (43.5)	**3 (15.0)**
≥500	408	30 (7.4)	10 (33.3)	**1 (10.0)**
Total	7918	564 (7.1)	290 (51.4)	111 (38.3)
p-value		0.005	0.056	0.012

Exploratory analyses (1)

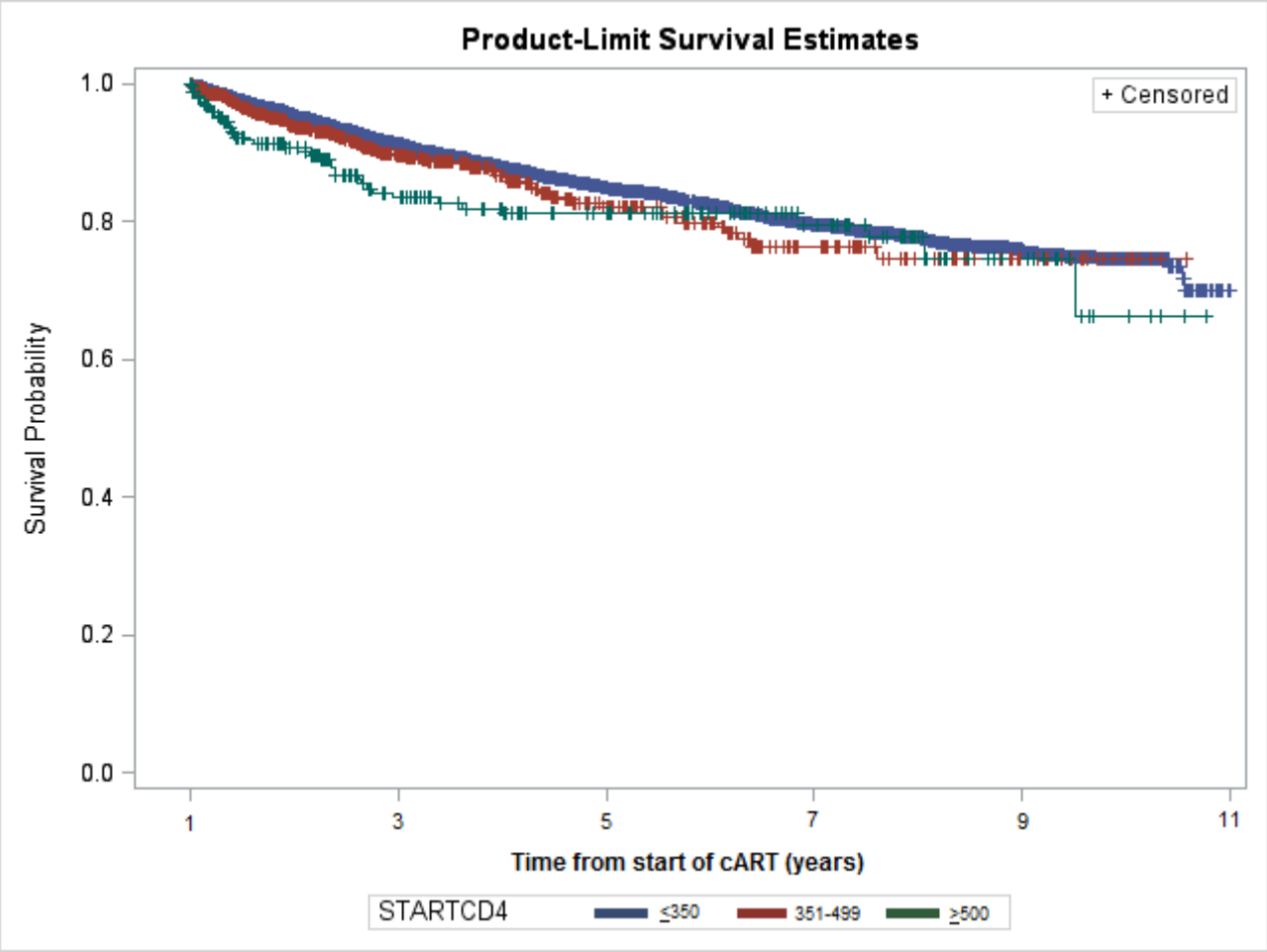
Aware of large limitation - reasons for starting cART with high CD4 counts not known

Undertook range of preliminary analyses to understand differences between 3 CD4 count groups

- Predictors of starting cART at high CD4 count
- How was CD4 count associated with following outcomes?
  - Virological suppression
  - Virological rebound
  - Treatment switching
  - Discontinuation of cART

Exploratory analyses (2)

Kaplan-Meier graph of time to discontinuation of cART, according to CD4 count at start of cART

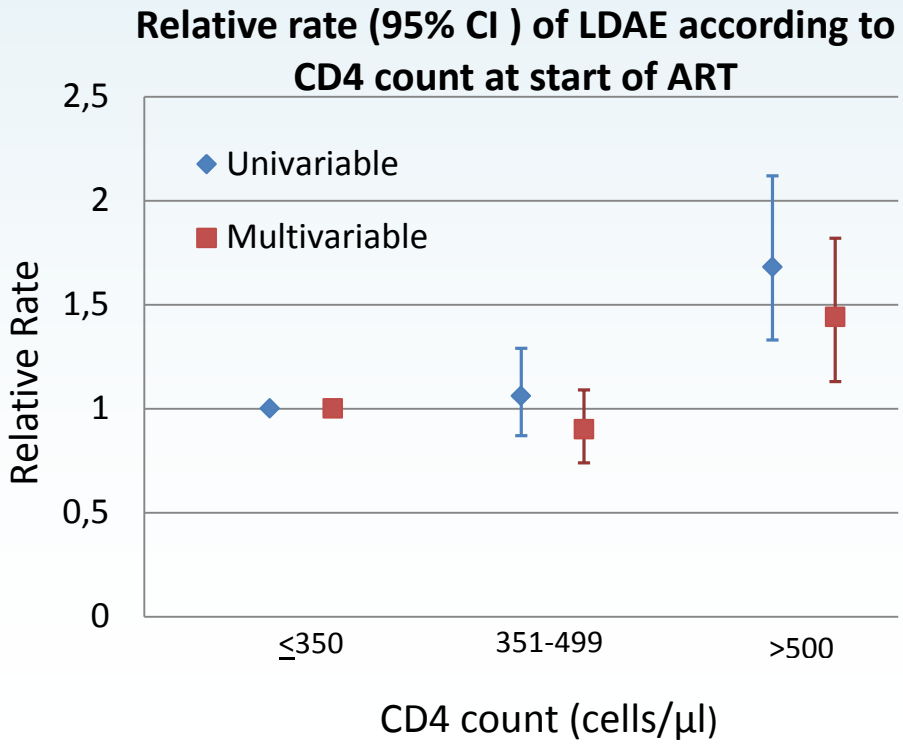


Main analyses (1)

- Should answer research question
- Provide estimates that are adjusted for measured confounders (regression models)
- Potential confounders:
  - Hepatitis B or Hepatitis C co-infection
  - ART regimen
  - Demographics (sex, ethnicity, exposure)

Main analyses (2)

- Should answer research question
- Provide estimates that are adjusted for measured confounders (regression models)

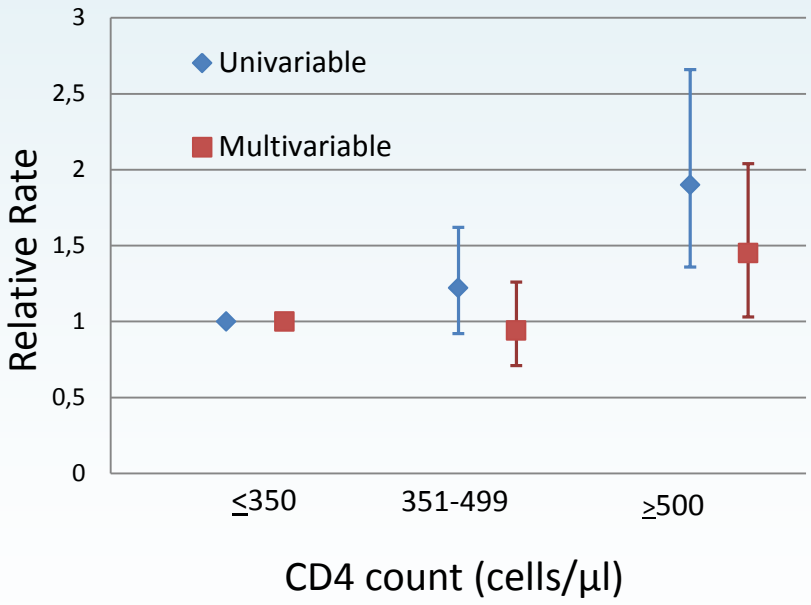


Sensitivity/Sub-group analyses (1)

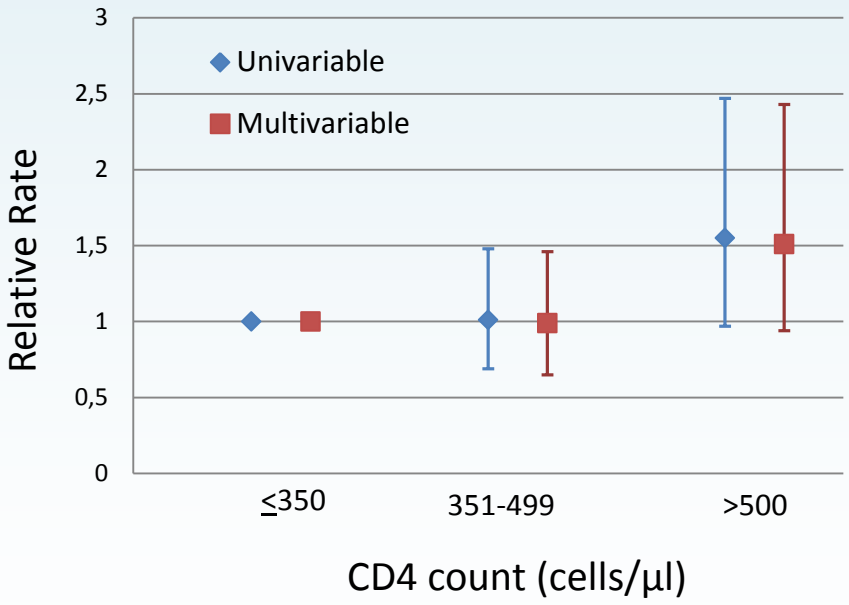
- If we change variable definitions, do our conclusions remain unchanged?
  - LDAE grouped by type (LFTS, renal function, blood, other) and analysed separately
  - Considered absolute change in laboratory measures
- If we change the population studied, do our conclusions remain unchanged?
  - Excluded those with HBV co-infection at cART start

Sensitivity/Sub-group analyses (2)

Relative rate (95% CI) of liver-related LDAE according to CD4 count at start of ART



Relative rate (95% CI) of blood-related LDAE according to CD4 count at start of ART



Write and publish the paper!

CONCISE COMMUNICATION

Laboratory adverse events and discontinuation of therapy according to CD4$^+$ cell count at the start of antiretroviral therapy

Sophie Jose[a], Killian Quinn[b], Teresa Hill[a], Clifford Leen[c], John Walsh[b], Phillip Hay[d], Martin Fisher[e], Frank Post[f], Mark Nelson[g], Mark Gompels[h], Margaret Johnson[i], David Chadwick[j], Richard Gilson[k], Caroline Sabin[a], Sarah Fidler[b], on behalf of the UK CHIC Steering Committee

Objective: Few data describe antiretroviral treatment (ART)-related adverse events when treatment is initiated at CD4 cell counts more than 350 cells/μl. We compared rates of laboratory-defined adverse events (LDAEs) according to CD4 cell count at ART initiation.

Design: Analysis of on-going cohort study.

Methods: ART-naive persons initiating ART from 2000 to 2010 were included. Chi-square, analysis of variance (ANOVA) and Kruskal–Wallis tests compared character-

Summary

- It is possible to answer a research question using data from your own clinic and without the need for big programme grants
- Need to be aware of potential biases and limitations that are present
- Lots of background work/analyses that don't make it to the finished article
- Get statistical advice from the start!