Developing an analysis plan

Outline of Session

- A real-life example
- Consideration of biases and limitations
- The stages of an analysis plan

Outline of Session

- A real-life example
- Consideration of biases and limitations
- The stages of an analysis plan

Example – UK CHIC Study

- Cohort study which collates data on HIV-positive individuals accessing care in the UK
- Data are those collected in routine HIV care: *CD4 counts, HIV viral load, ART, basic demographics*
- Because data are collected in advance, important to be aware of potential limitations and be pragmatic
- UK CHIC was set up and approved for research purposes and data are pseudonomysed so informed consent not required

The idea

- Approached by clinician at a London-based HIV clinic
- Interested in whether treatment toxicities were more likely to occur in those starting cART with high CD4 counts
- Secondary end-point HPTN 052
- START yet to report
- Would it be possible to investigate this using the routine clinic data collected through UK CHIC?

The Research Question

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

Outline of Session

- A real-life example
- Consideration of biases and limitations
- The stages of an analysis plan

Before you plan an analysis

- Must fully consider potential biases and limitations of your study including:

Before you plan an analysis

- Must fully consider potential biases and limitations of your study including:
  - Confounding
  - Missing data
  - Attrition bias
  - Observer bias
  - Survivorship bias
  - Lead-time bias
  - Using routine data

Before you plan an analysis

- Must fully consider potential biases and limitations of your study including:
  - Confounding
  - Missing data
  - Attrition bias
  - Observer bias
  - Survivorship bias
  - Lead-time bias
  - Using routine data

Before you plan an analysis

Confounding

- Who starts cART with high CD4 counts in routine practice?
- Likely to differ from those starting cART below 350 cells/mm^3 threshold
- Do we collect data on important confounders?
- Can we adjust for these in our analyses?
- Unmeasured confounding

What biases/limitations do we need to consider?

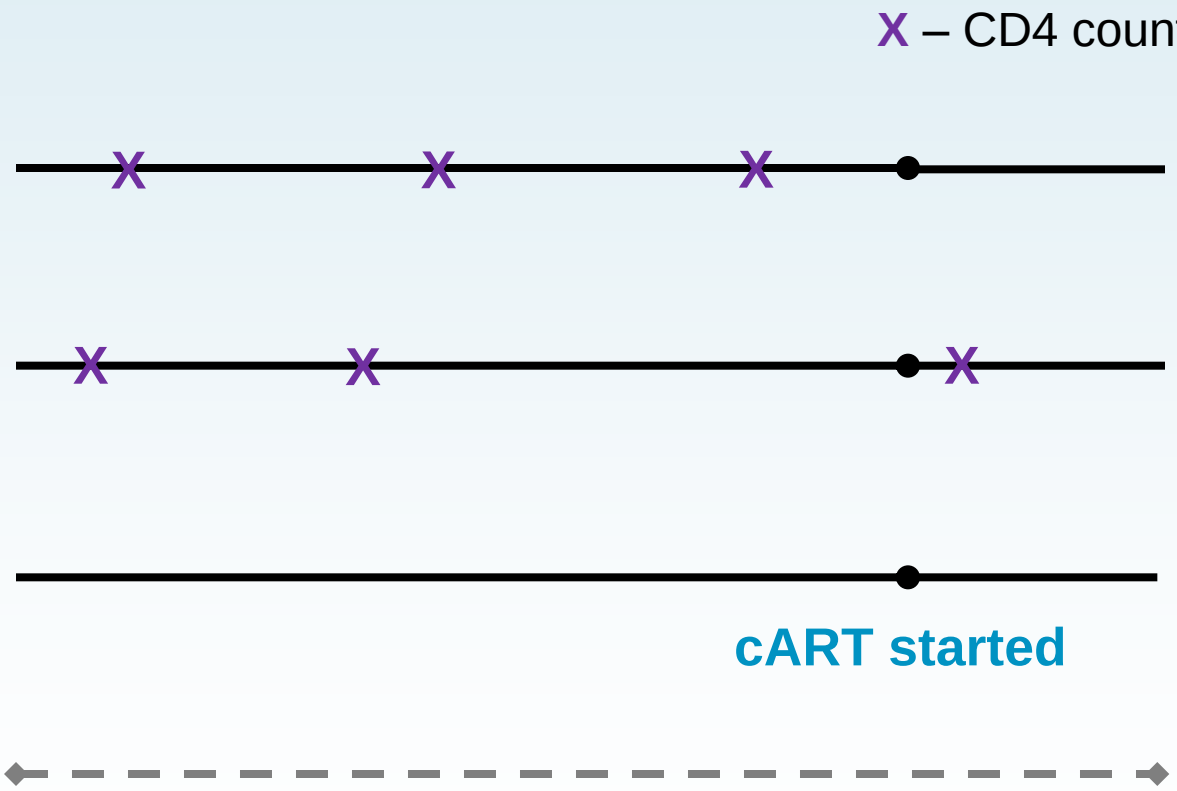
		CD4 count at start of ART (cells/mm³)		
		≤350	351-499	≥500
Sex, n (%)	**Male**	**6147 (78.2)**	**958 (87.2)**	**406 (90.8)**
Ethnicity, n (%)	**White**	**4586 (58.4)**	**787 (71.6)**	**334 (74.7)**
	Black African	1861 (23.7)	131 (11.9)	42 (9.4)
	Black other	404 (5.1)	52 (4.7)	17 (3.8)
	Other/unknown	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)

What biases/limitations do we need to consider?

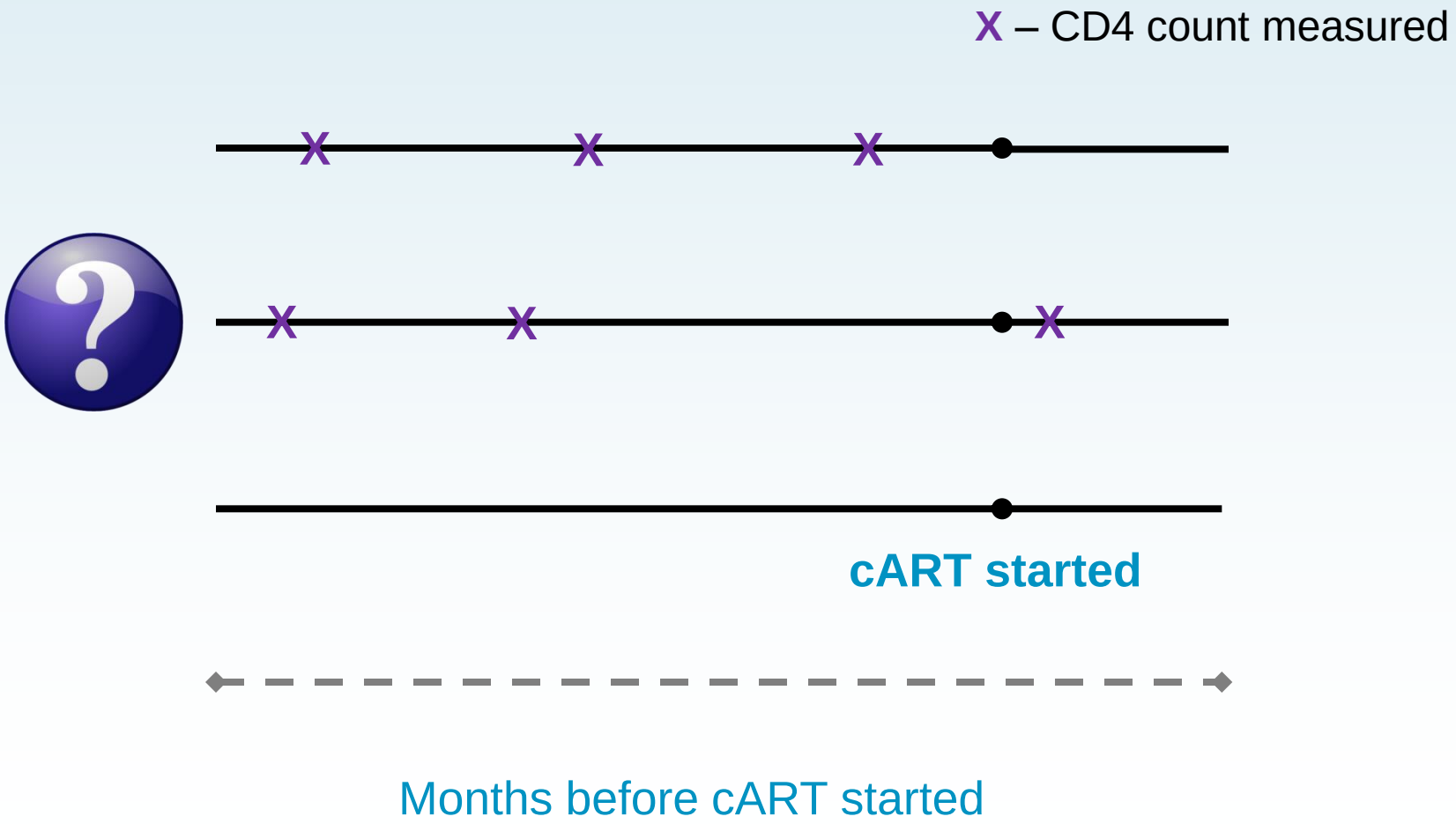
Missing data

- Need CD4 count at cART start
- Censor patient follow-up when lost to follow-up

What biases/limitations do we need to consider?



What biases/limitations do we need to consider?

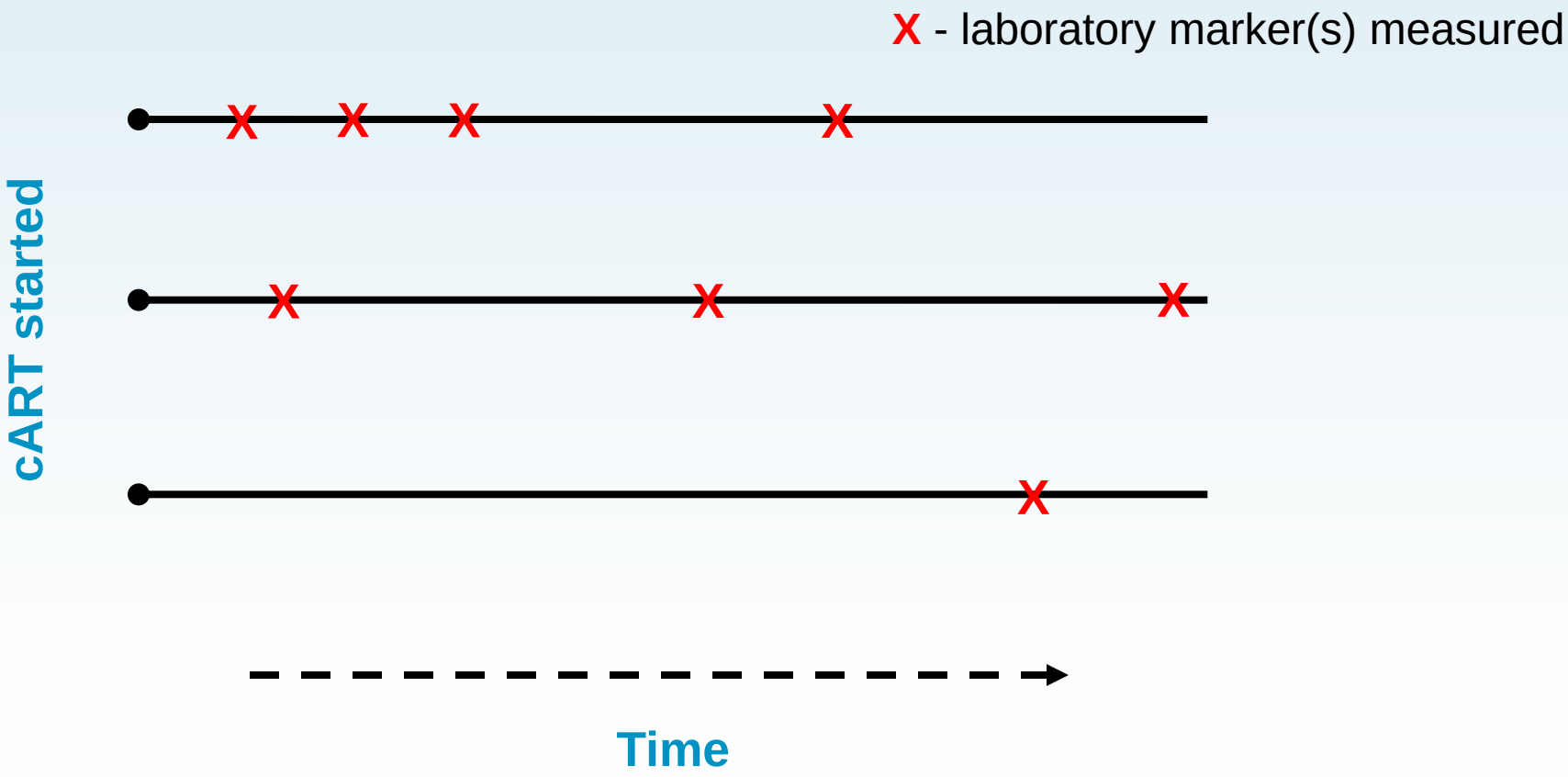


What biases/limitations do we need to consider?

Infrequent monitoring

- Laboratory tests not performed at regular intervals in all patients
- More likely to have a test if sick or displaying symptoms

What biases/limitations do we need to consider?



Outline of Session

- A real-life example
- Consideration of biases and limitations
- The stages of an analysis plan

Stages of an analysis plan

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?

Stages of an analysis plan

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

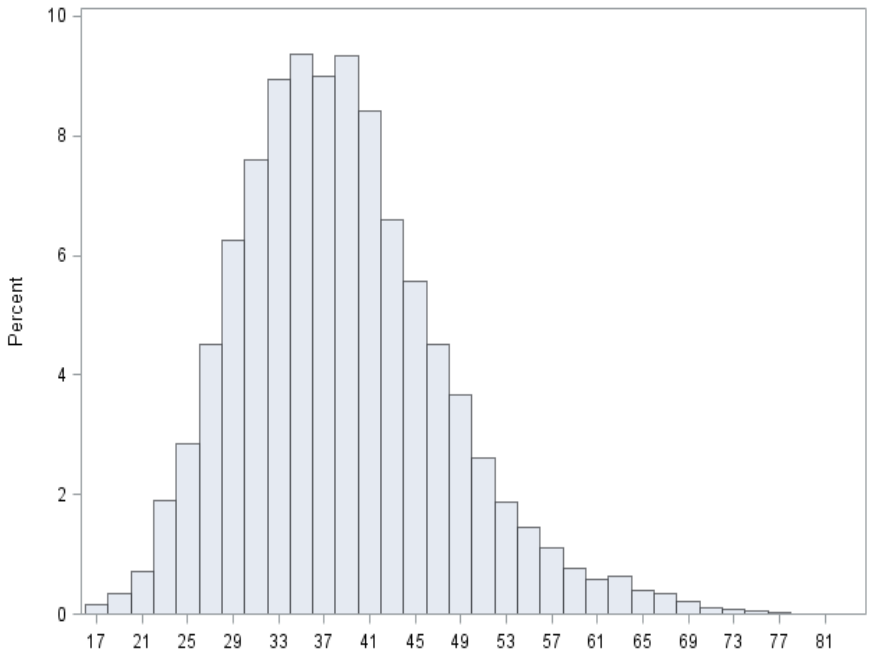
- Are methods valid?

Descriptive Analyses

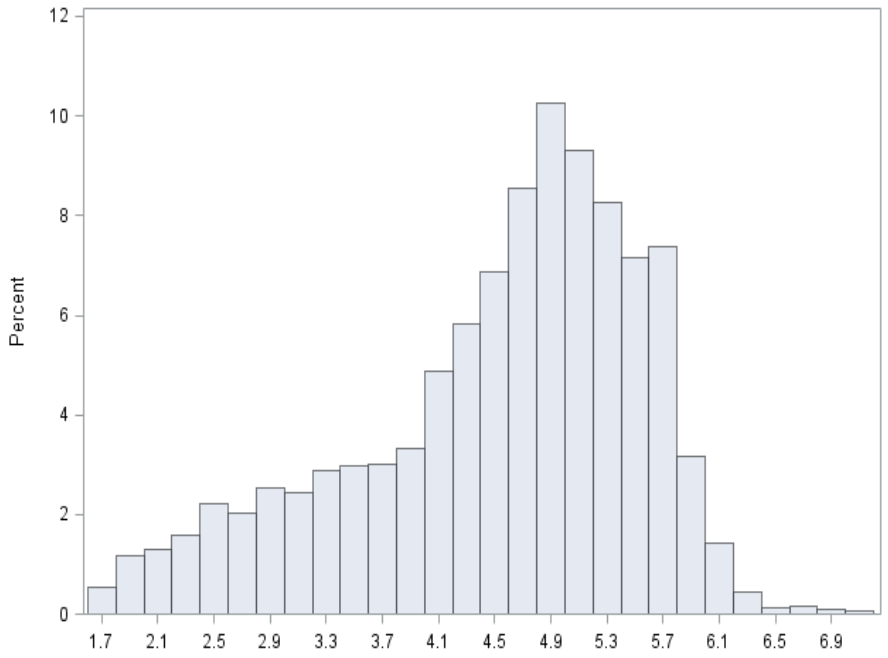
1. Data checks
2. Patient flow through study
3. Number of people with and without exposure of interest
4. Describe baseline demographic and clinical characteristics
   - According to exposure of interest
   - Appropriate univariate tests
5. Study follow-up
6. Outcome

Descriptive Analyses

Data checks



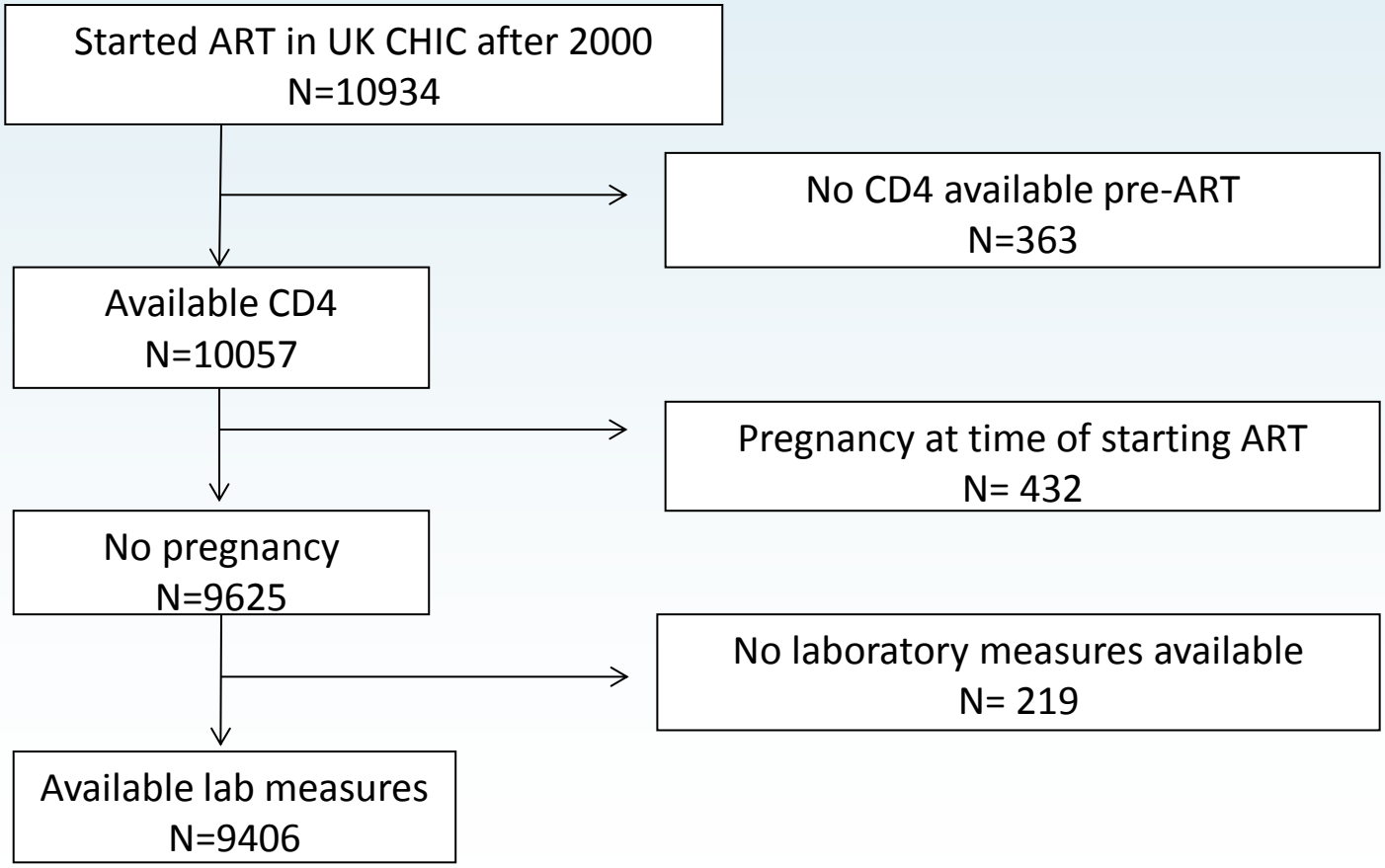
Age at cART start



Log_{10} VL at cART start

Descriptive Analyses

Patient flow-chart through study



Descriptive Analyses

Baseline characteristics

		CD4 count at start of ART (cells/mm^3)			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

Patient follow-up

		CD4 count at start of ART (cells/mm^3)		
		≤350	351-499	≥500
Patient follow up, years	Sum	24379.8	2318.6	922.2
	Median (IQR)	2.5 (0.0, 11)	1.4 (0.0, 10.6)	1.0 (0.0, 10.6)
Average number of laboratory tests/ year	Median (IQR)	2 (0,3)	2 (0,3)	3 (0,5)

Descriptive Analyses

Outcome

Baseline CD4 count, cells/mm^3	N	Number of LDAE	%	Rate /100 person years
≤350	7860	1094	13.9	4.5 (4.2, 4.8)
351-499	1099	113	10.3	4.9 (4.0, 5.8)
≥500	447	76	17.0	8.2 (6.4, 10.1)

Stages of an analysis plan

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

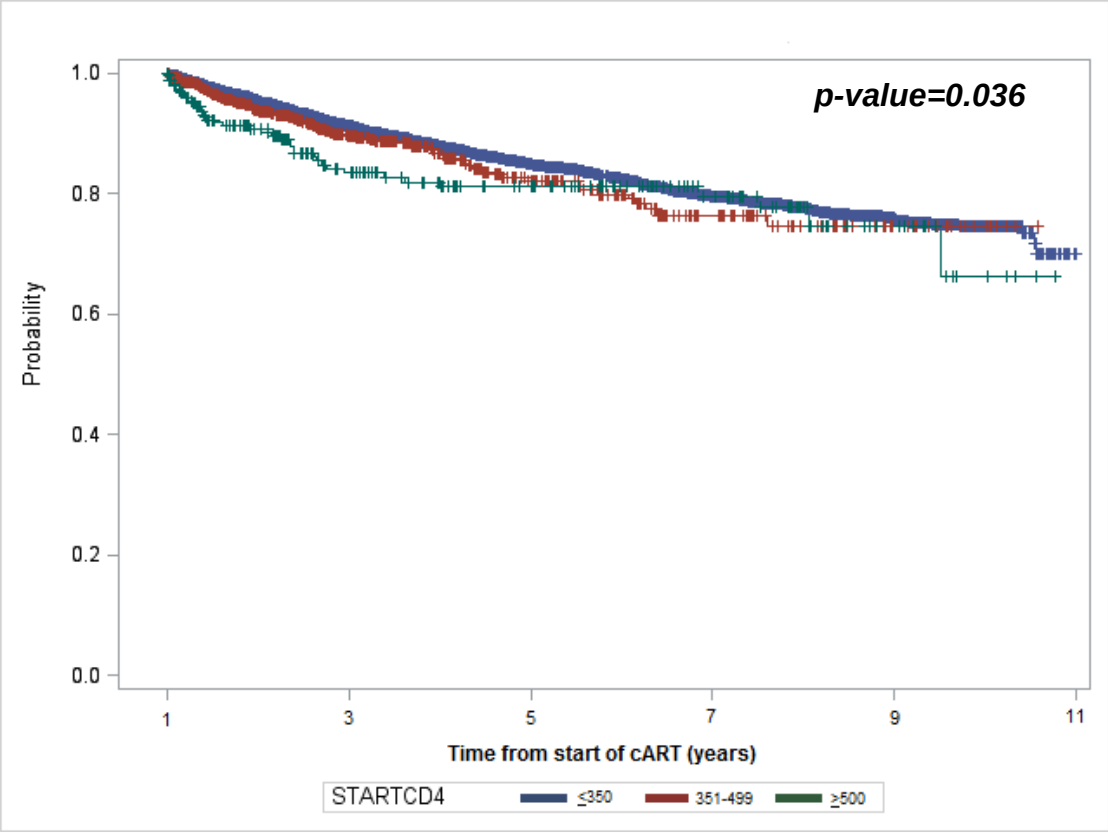
- Are methods valid?

Exploratory analyses

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

Exploratory analyses

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



Stages of an analysis plan

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

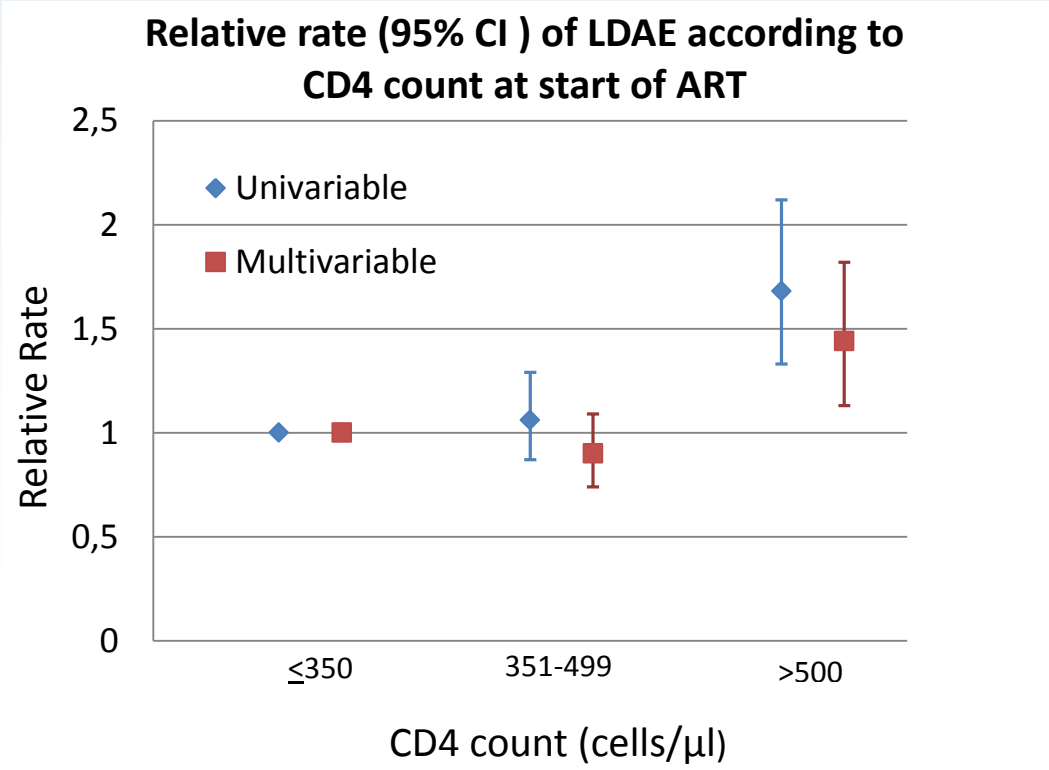
- Are methods valid?

Main analysis

- 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 analysis

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



Stages of an analysis plan

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?

Sensitivity/sub-group analyses

If we change variable definitions, do our conclusions remain unchanged?

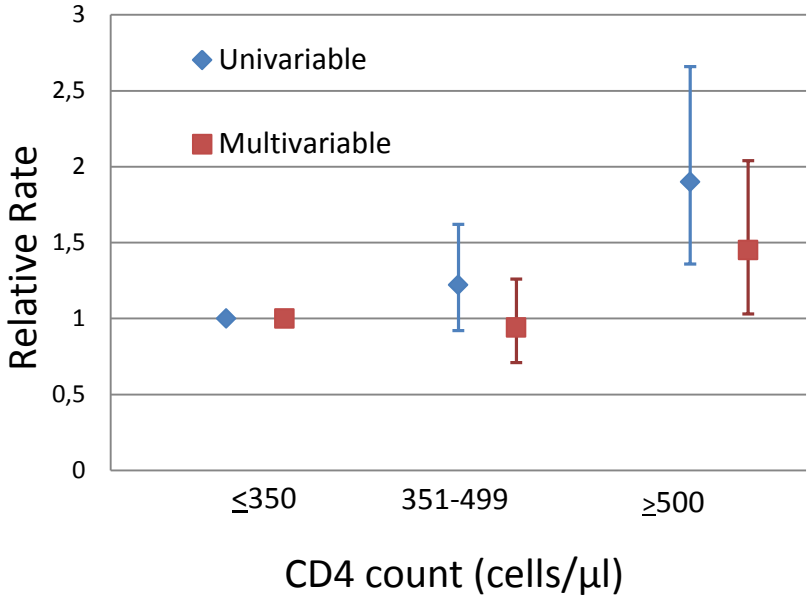
- LDAE grouped by type (LFTS, renal function, blood, other) and analysed separately
- Considered absolute change in laboratory measures

If we change the population studied, do our conclusions remain unchanged?

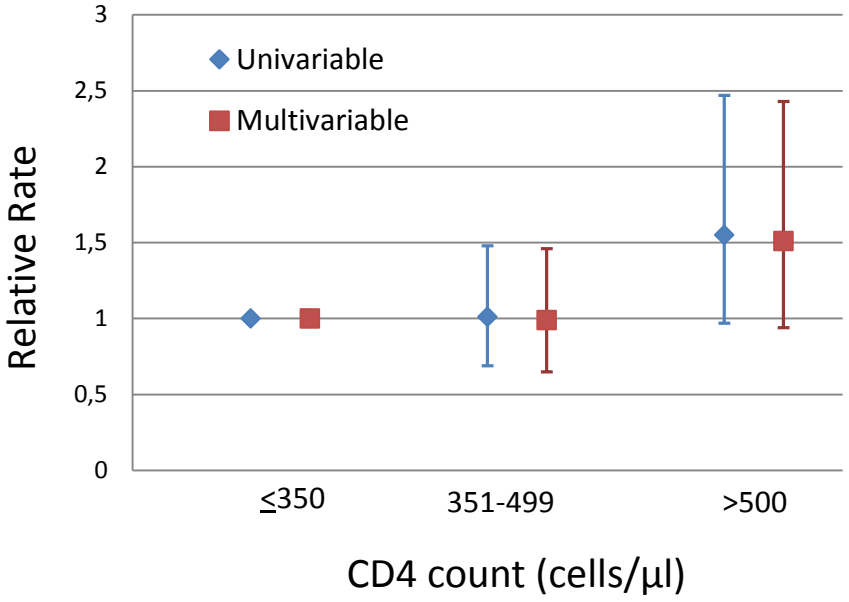
- Excluded those with HBV co-infection at cART start

Sensitivity/sub-group analyses

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



Summary

- It is important to consider potential biases and limitations of your study before planning an analysis
- An analysis should be planned in advance of conducting the study
- There are four main stages of an analysis plan: descriptive, exploratory, main and sub-group and sensitivity analyses
- Even if you aren’t conducting analyses, important to get to know your data and understand analysis plan