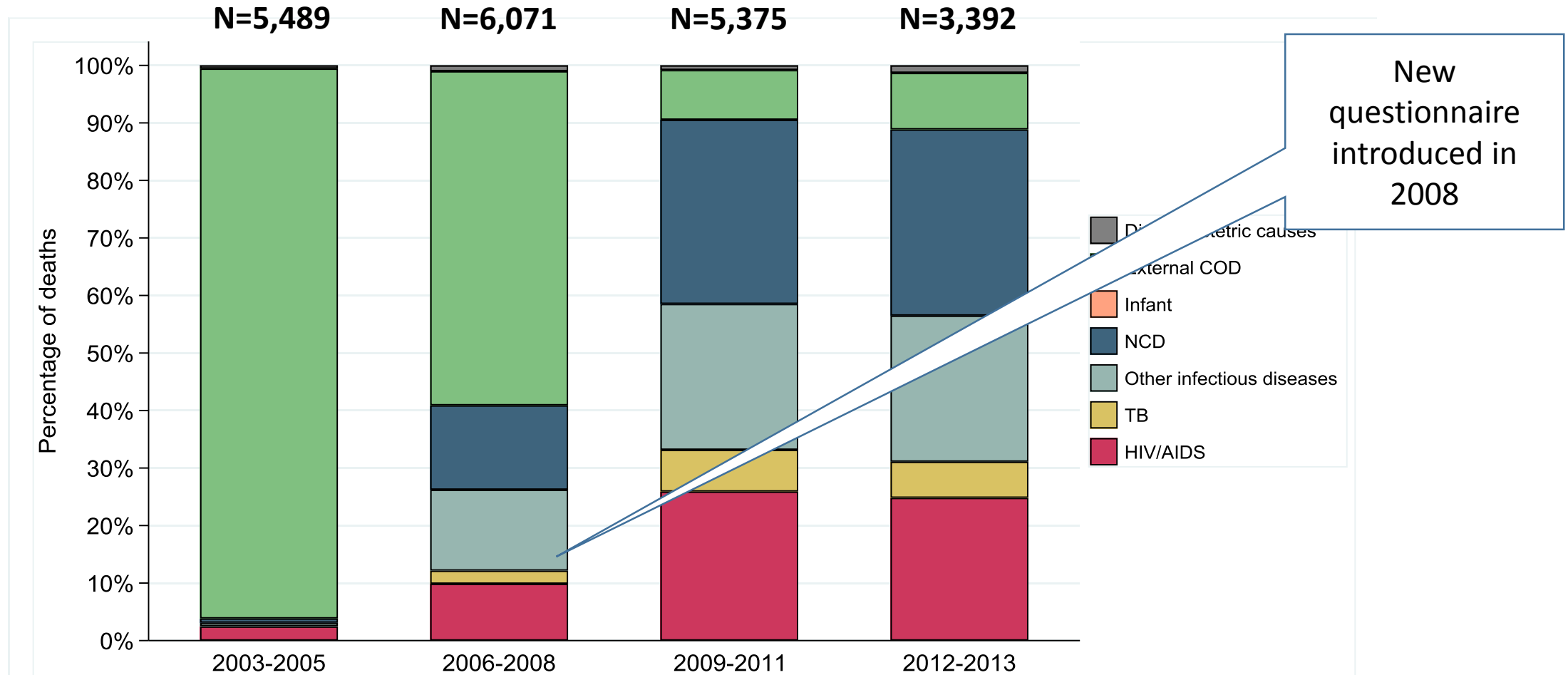# Exploratory analysis based on openVA output

November 12, 2018

"We are finally at the last step...except, it is not our last step."

Are the results plausible?
Example 1: Cause-specific mortality fractions by year

Graph: Cause-specific mortality fractions in adults, by calendar year

Are the results plausible?
Example 1: Cause-specific mortality fractions by year

Table: Response to VA indicator on road traffic accident reported in verbal autopsy for each death by year in which the verbal autopsy was conducted

Road traffic accident reported in VA	Year of death			
	2003-2005	2006-2008	2009-2011	2012-2013
Yes	5,219	3,323	87	73
No	270	212	4,298	2,417
Missing	0	2,536	990	902

Are the results plausible?
Example 1: Cause-specific mortality fractions by year

Table: Percentage of deaths with road traffic accident reported in verbal autopsy by year in which the verbal autopsy was conducted

Road traffic accident reported in VA	Year of death			
	2003-2005	2006-2008	2009-2011	2012-2013
Yes	95.1	54.7	1.6	2.2
No	4.9	3.5	80.0	71.3
Missing	0	41.8	18.4	26.6

There was a coding error for the question on traffic accidents from 2003-2008 - those who did not have a road traffic accident were coded as having a road traffic accident.
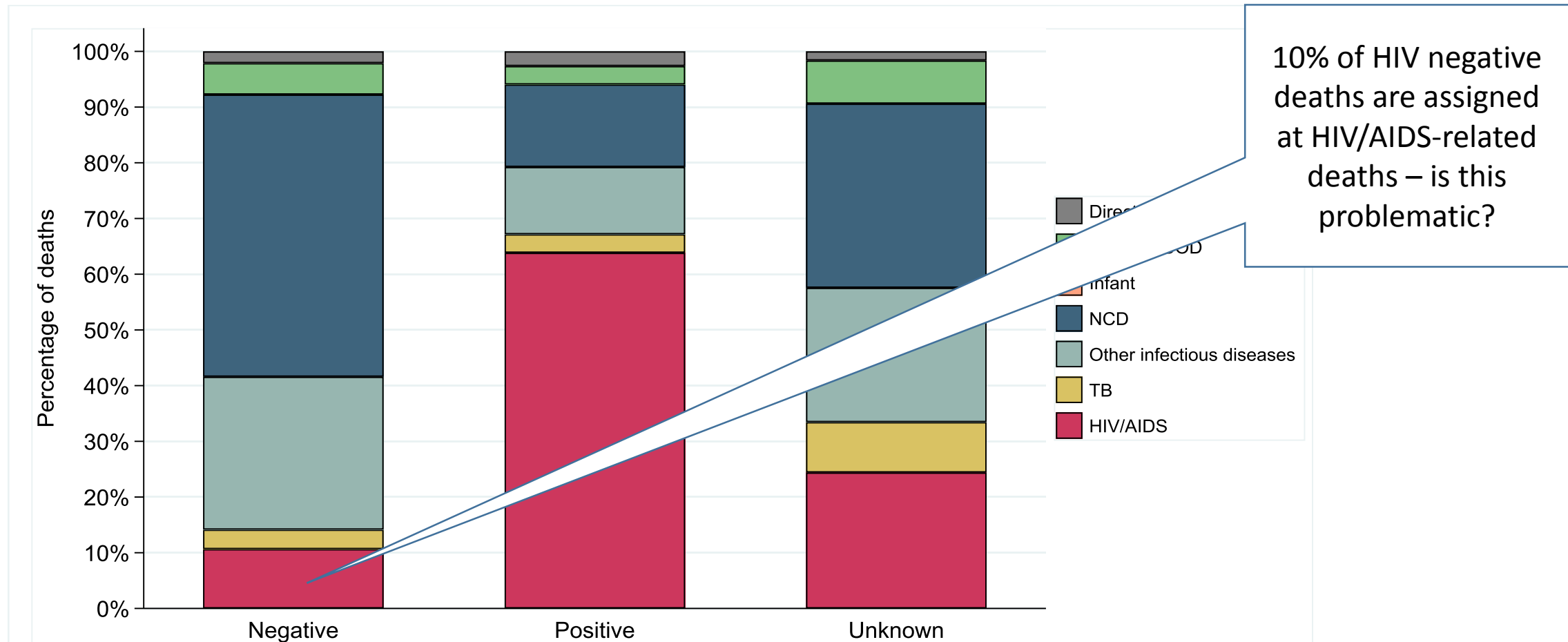
Are the results plausible?
Example 1: Cause-specific mortality fractions by year

Graph: Cause-specific mortality fractions in adults, by calendar year

Are the results plausible?
Example 2: Cause-specific mortality fractions by HIV status

Graph: Cause-specific mortality fractions in adults, by HIV status

Are the results plausible?
Example 2: Cause-specific mortality fractions by HIV status

Graph: Cause-specific mortality fractions in adults, by HIV status (comparison site)
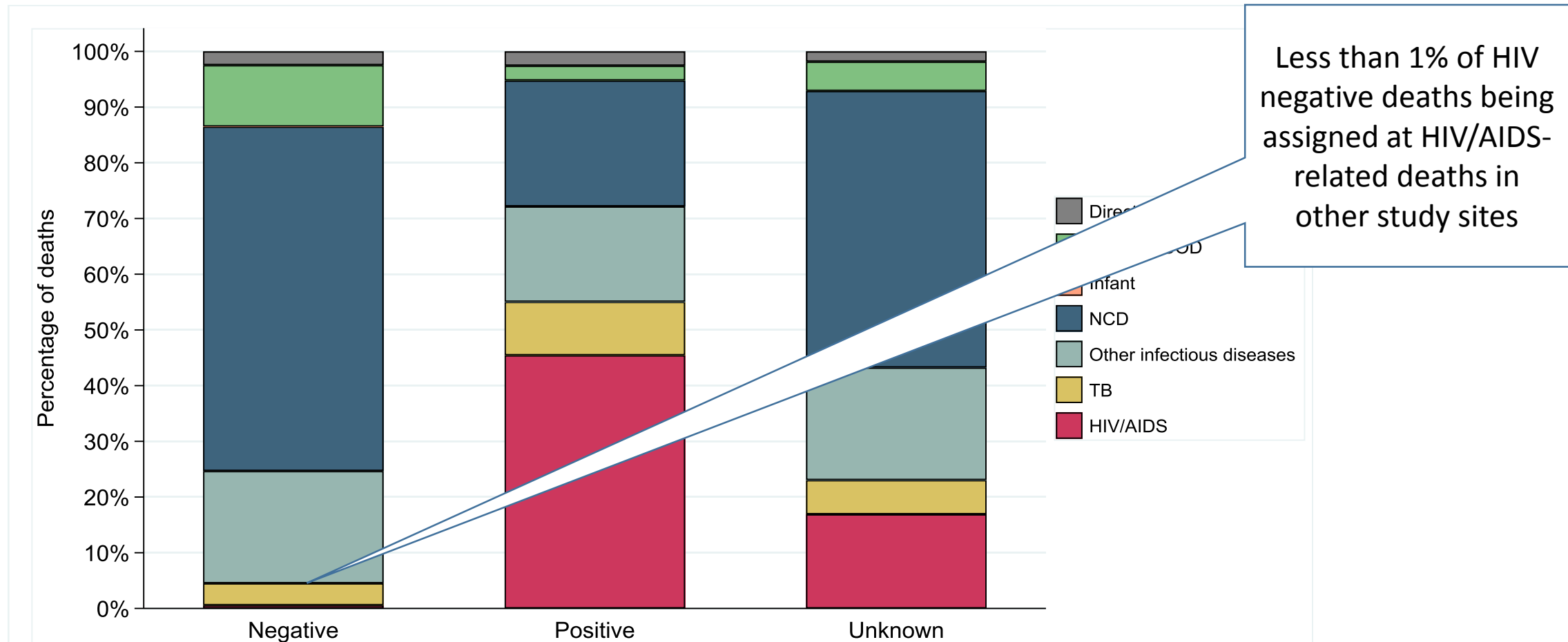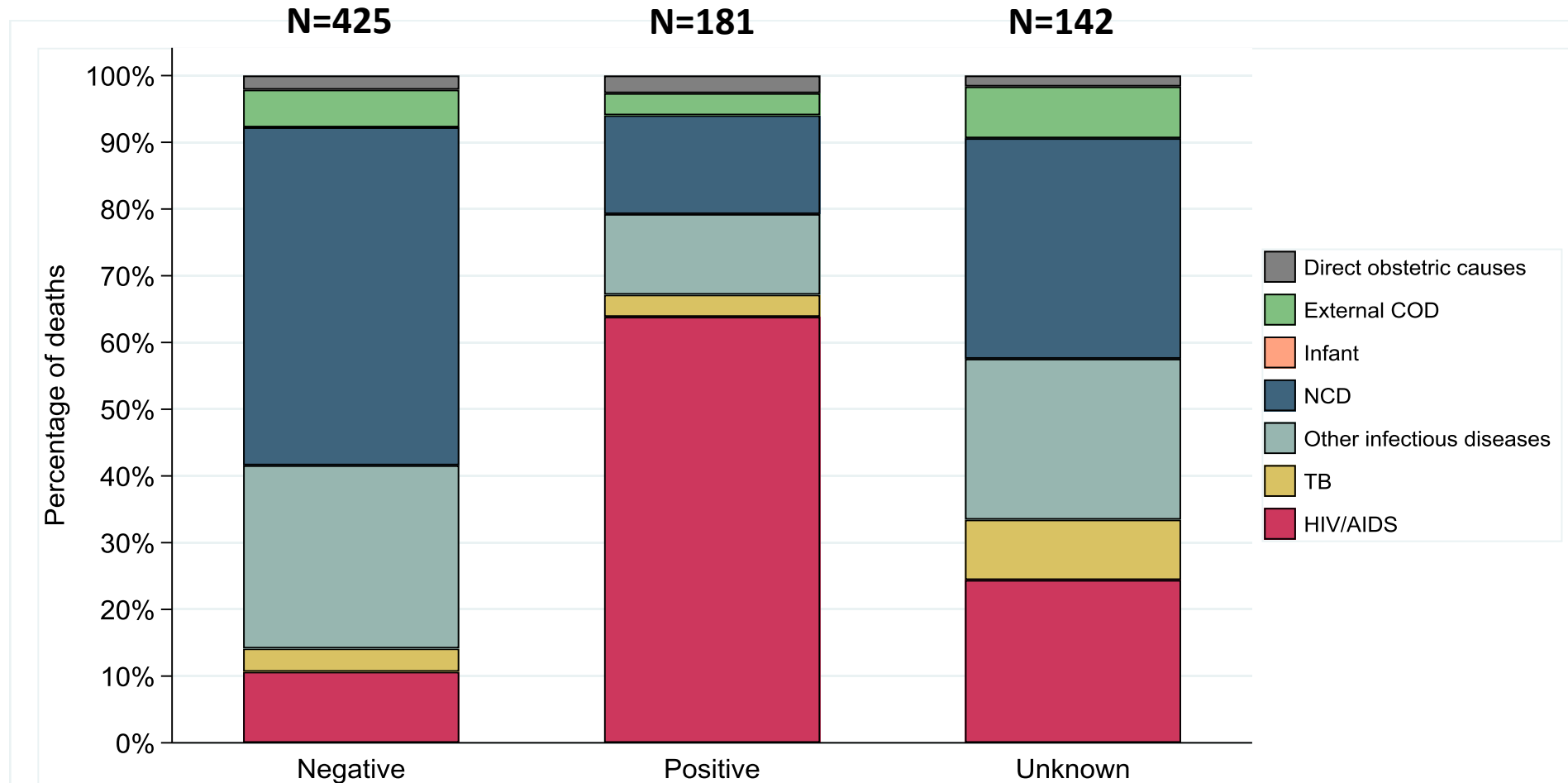
Are the results plausible?
Example 2: Cause-specific mortality fractions by HIV status

Graph: Cause-specific mortality fractions in adults, by HIV status

Are the results plausible?
Example 2: Cause-specific mortality fractions by HIV status

Table: Percentage of deaths with classic HIV/AIDS symptoms reported in verbal autopsy by HIV status

	HIV positive	HIV negative	HIV status unknown
Weight loss	61.3	42.6	54.9
			71.3
			N/A
			15.5
			29.6
			15.5

In this example, we have information on HIV status from an independent source. If you are finding an unexpected prevalence of HIV amongst deaths in your total study population, how could you explore whether there might be errors in the frequency of some key HIV/AIDS symptoms?

High prevalence of symptoms amongst HIV negative -> cannot correct at analysis stage, but need to consider what is happening in VA interviews.
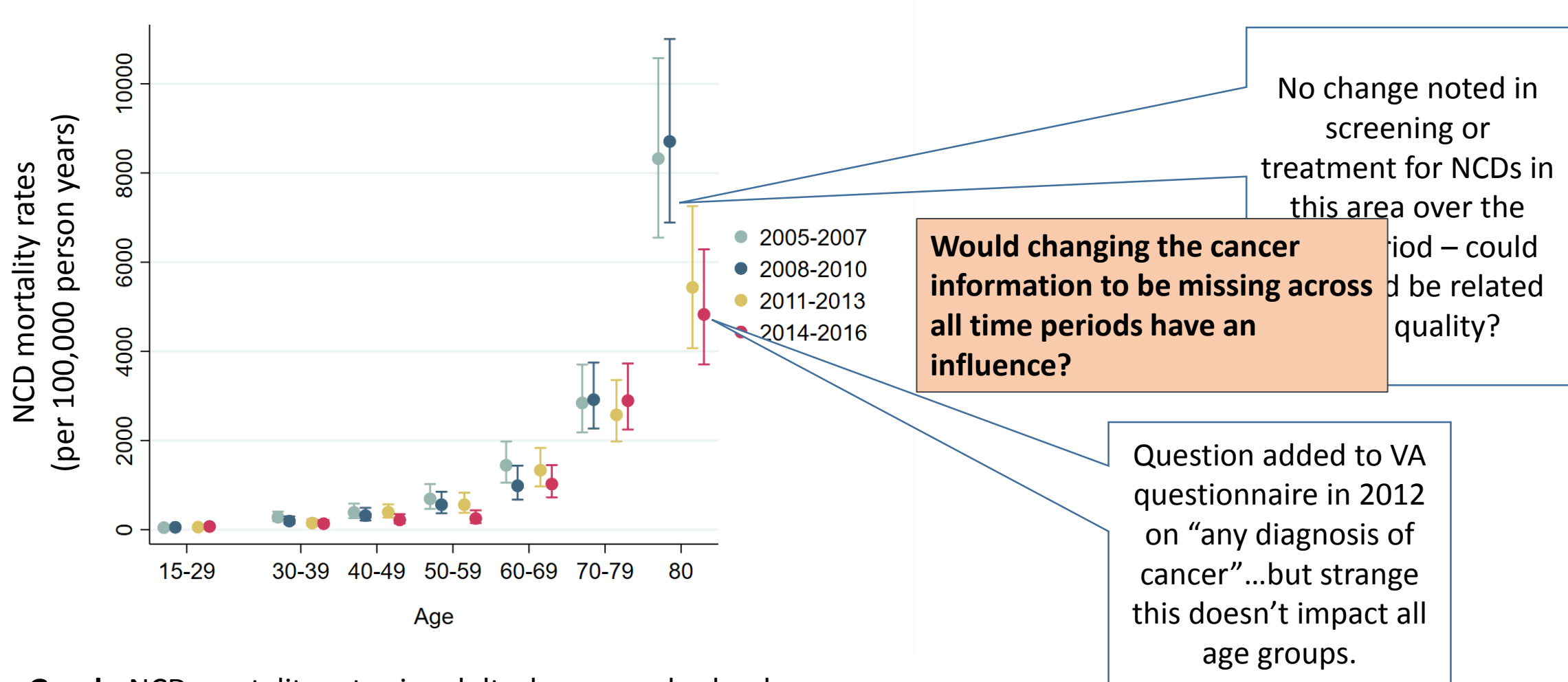
Are the results plausible?
Example 2: Cause-specific mortality fractions by HIV status

Table: Percentage of deaths with classic HIV/AIDS symptoms reported in verbal autopsy

	% of total population	% in comparison population with similar HIV prevalence
Weight loss	49.5	48.8
Diarrhoea	69.9	24.3
Acute respiratory tract infection	13.0	2.0
Oral Candidiasis	N/A	9.9
Abscesses/sores	14.6	12.2
Wasting	22.9	9.0
Recent TB	15.9	10.2

Are the results plausible?
Example 3: Non communicable disease mortality rates over time

Graph: NCD mortality rates in adults, by age and calendar year

Are the results plausible?
Example 3: Non communicable disease mortality rates over time

Graph: NCD mortality rates in adults, by age and calendar year
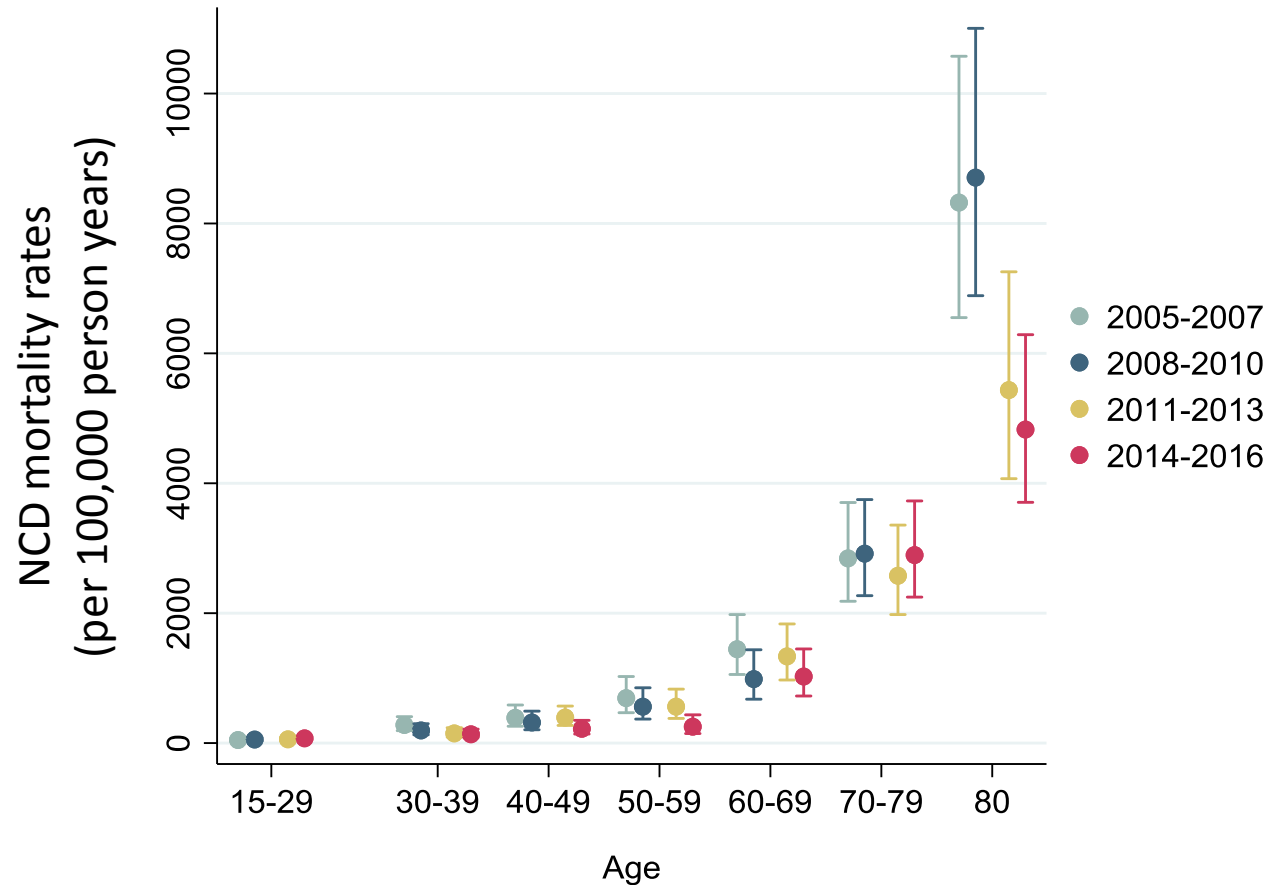
Would changing the cancer information to be missing across all time periods have an influence?
….NO

What other data can we use to assess plausibility?

We can assess the **plausibility of our cause-specific mortality fractions** (CSMFs) by looking at:

- Comparison of CSMFs from different algorithms

- Distribution of causes of death by basic demographic characteristics and changes over time

- Symptom patterns for certain causes of death

Sometimes, we can draw on **additional data sources** to help assess the quality of **cause of death assignments** – **can you think of any of these?**

Plausibility of CSMFs

➢Physician review

➢Data from morbidity surveys

➢Clinical record data

➢Minimally invasive autopsies

➢Conventional autopsy

Plausibility of CSMFs

➢**Physician review**

➢Data from morbidity surveys

➢Clinical record data

➢Minimally invasive autopsies

➢Conventional autopsy

Physician review

Verbal autopsies are reviewed by two physicians, with a third physician consulted in cases where there is disagreement

Discuss the advantages and disadvantages of physician review compared with algorithms?

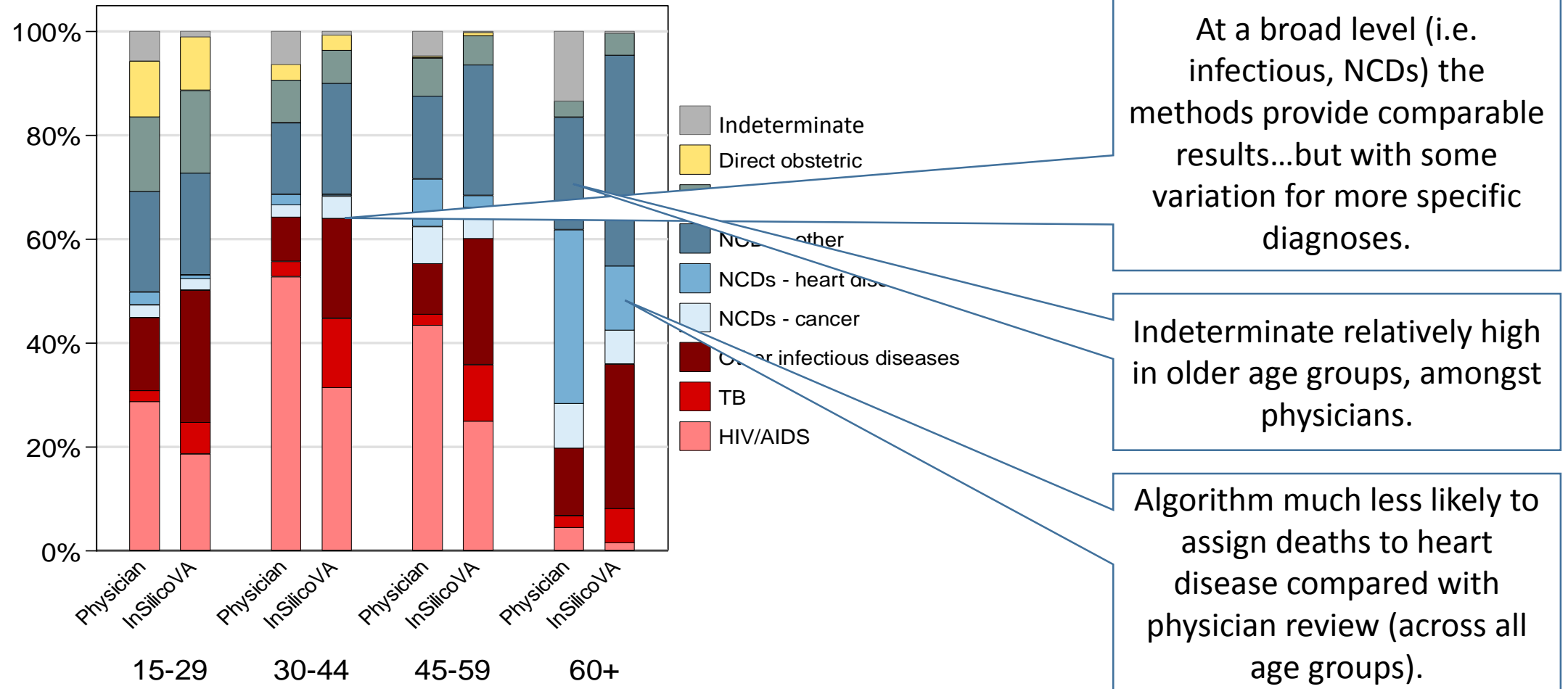| Advantages	Disadvantages
• Often coded with underlying, contributory and direct causes and in line with ICD codes • Make use of information captured in the narrative	• Expensive and time consuming for clinicians • Poor inter-observer agreement • Inconsistencies between different areas

Example: Agreement of InSilicoVA and physician review (population level)
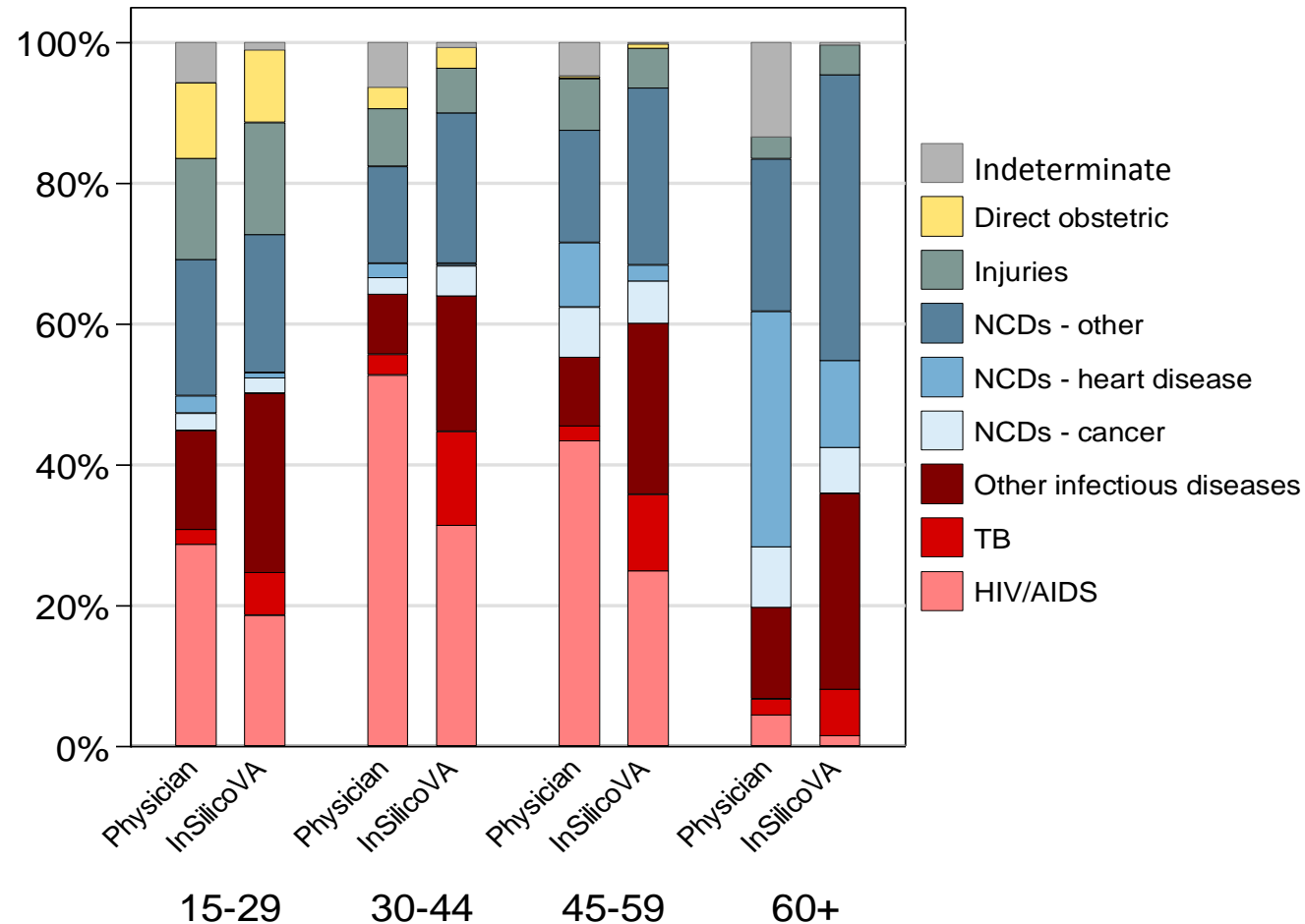
What conclusions would you draw on the basis of this graph?

Graph: Cause-specific mortality fractions in adults, by age group and method of assigning cause of death

Example: Agreement of InSilicoVA and physician review (population level)

At a broad level (i.e. infectious, NCDs) the methods provide comparable results…but with some variation for more specific diagnoses.

Indeterminate relatively high in older age groups, amongst physicians.

Algorithm much less likely to assign deaths to heart disease compared with physician review (across all age groups).

Graph: Cause-specific mortality fractions in adults, by age group and method of assigning cause of death

Example: Agreement of InSilicoVA and physician review (individual level)

Table: Number of deaths by cause assigned by physician review and cause assigned by InSilicoVA

InSilicoVA main cause of death	Physician main cause of death							
	HIV/AIDS	TB	Other communicable	NCD	External	Direct obstetric	Unknown	TOTAL
HIV/AIDS	294	3	6	22	0	0	7	332
TB	96	30	8	50	0	1	7	192
Other communicable	82	6	167	196	10	5	62	528
NCD	90	12	62	592	22	4	109	891
External	3	0	2	17	109	0	9	140
Direct obstetric	3	0	0	4	0	38	2	47
Indeterminate	4	0	1	6	0	0	0	11
TOTAL	572	51	246	887	141	48	196	2141

What conclusions would you draw on the basis of this table?

Example: Agreement of InSilicoVA and physician review (individual level)

Of the 572 deaths assigned as HIV by physicians, only 294 deaths had the most likely cause as HIV/AIDS.

Table: Number of deaths by cause assigned by physician review and cause assigned by InSilicoVA

Better agreement on other broader cause categories.

InSilicoVA main cause of death	Physician main cause of death							
	HIV/AIDS	TB	Other communicable	NCD	External	Direct obstetric		TOTAL
HIV/AIDS	294	3	6	22	0	0		332
TB	96	30	8	50	0	1	7	192
Other communicable	82	6	167	196	10	5	62	528
NCD	90	12		592	22	4	109	891
External	3	0	2	17	109		9	140
Direct obstetric	3	0	0	4	0	8	2	47
Indeterminate	4	0	1	6	0	0	0	11
TOTAL	572	51	246	887	141	48	196	2141

Take-away message? Useful at the population level, but not at individual level…

Conclusions:

➢Checking the plausibility of the cause-specific mortality fractions (CSMF) is critical, as errors can be introduced during any one of the many steps involved in VA.

➢There are a number of ways to explore suspicious findings:
  ➢Check the coding of input data relating to that cause of death *(Example 1: injury deaths)*
  ➢Check the prevalence of certain symptoms in your population with those in another population *(Example 2: HIV/AIDS deaths)*
  ➢Look at the influence a particular input might be having on the CSMFs, by rerunning the data leaving that particular input blank *(Example 3: NCD deaths)*

➢*NOTE:* Many of the issues that have arisen across study sites are related to changes in questionnaires over time and/or small sample sizes.