



Traceability and AI for Better Understanding, Communication, and QC of Trials



Tomás Sabat Stöf sel

CEO & Co-Founder, Verisian

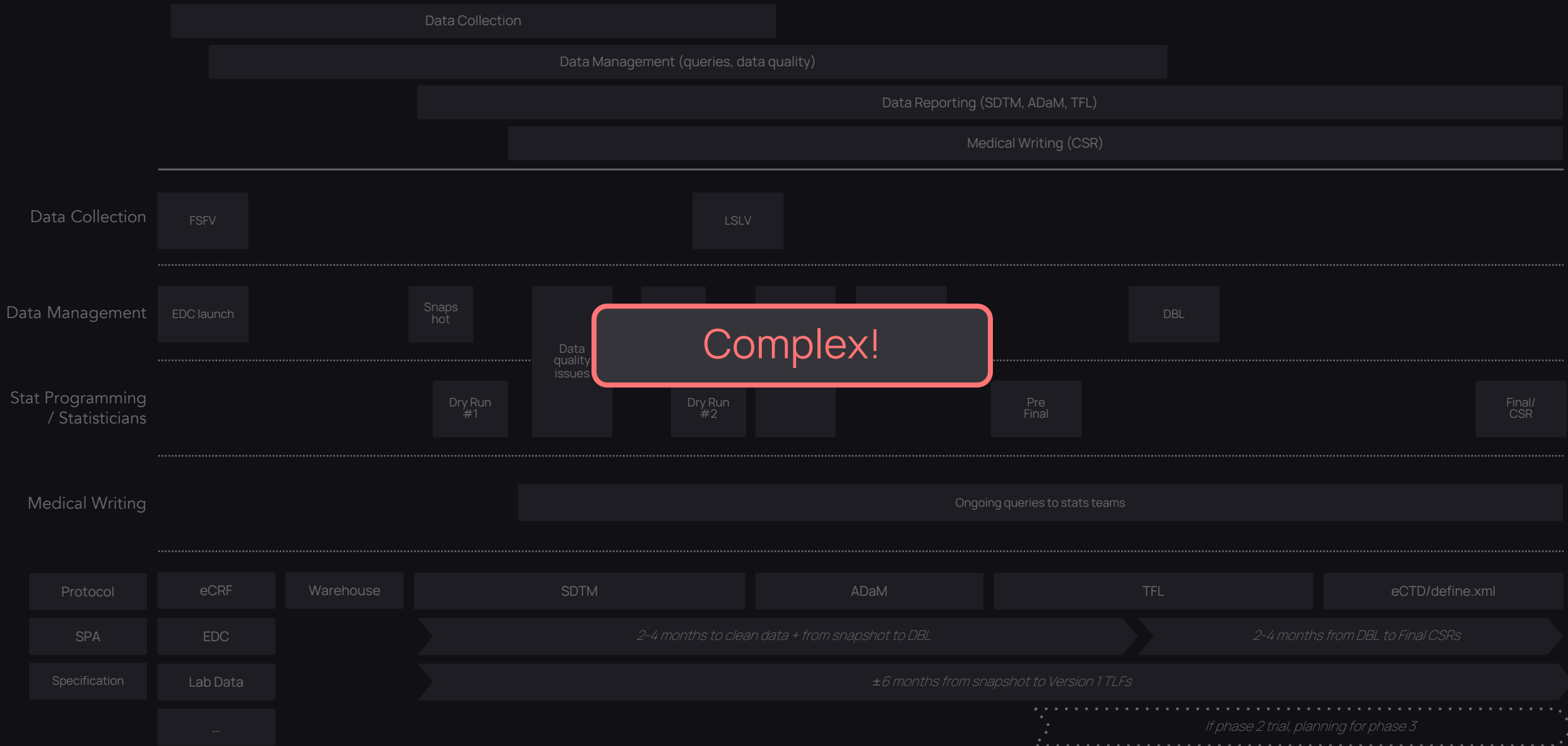
High attrition rates, high costs, and long development times
call for more agile processing of data,
to quantify the risks and facilitate timely decisions

Leading to more agile drug development

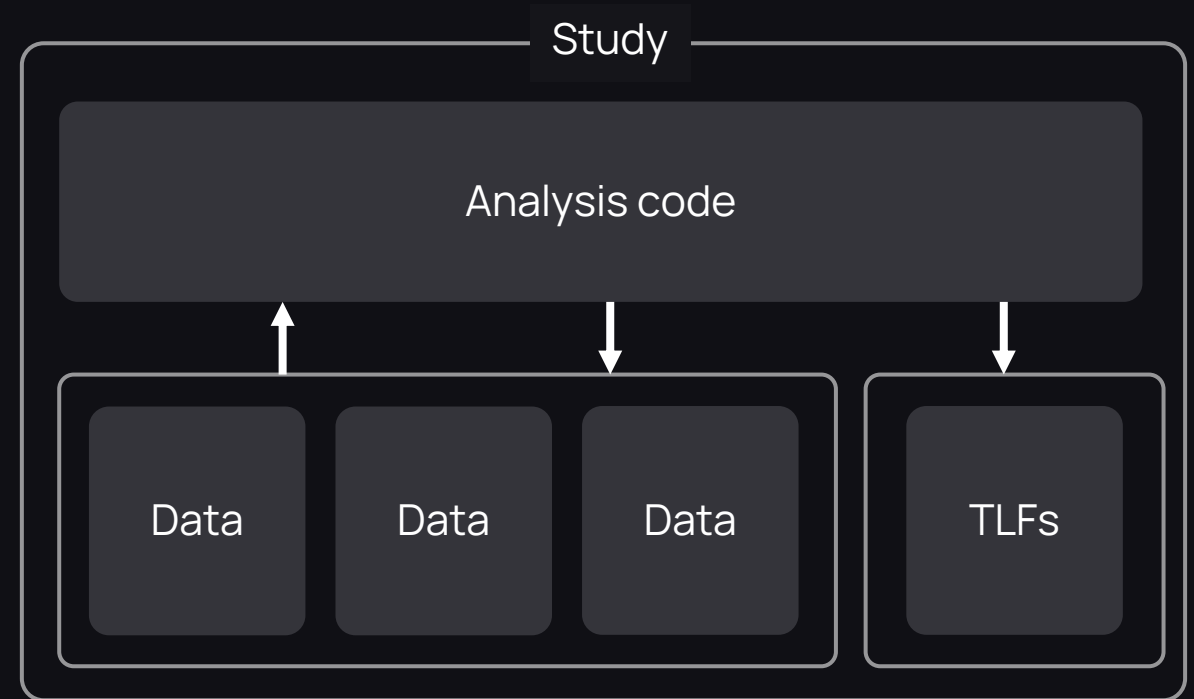
The availability of data
and the ability to derive action from data
will determine the **success of drug development programs**

Why can't we release a drug immediately after the last pivotal trial?

Drug submission data workflow

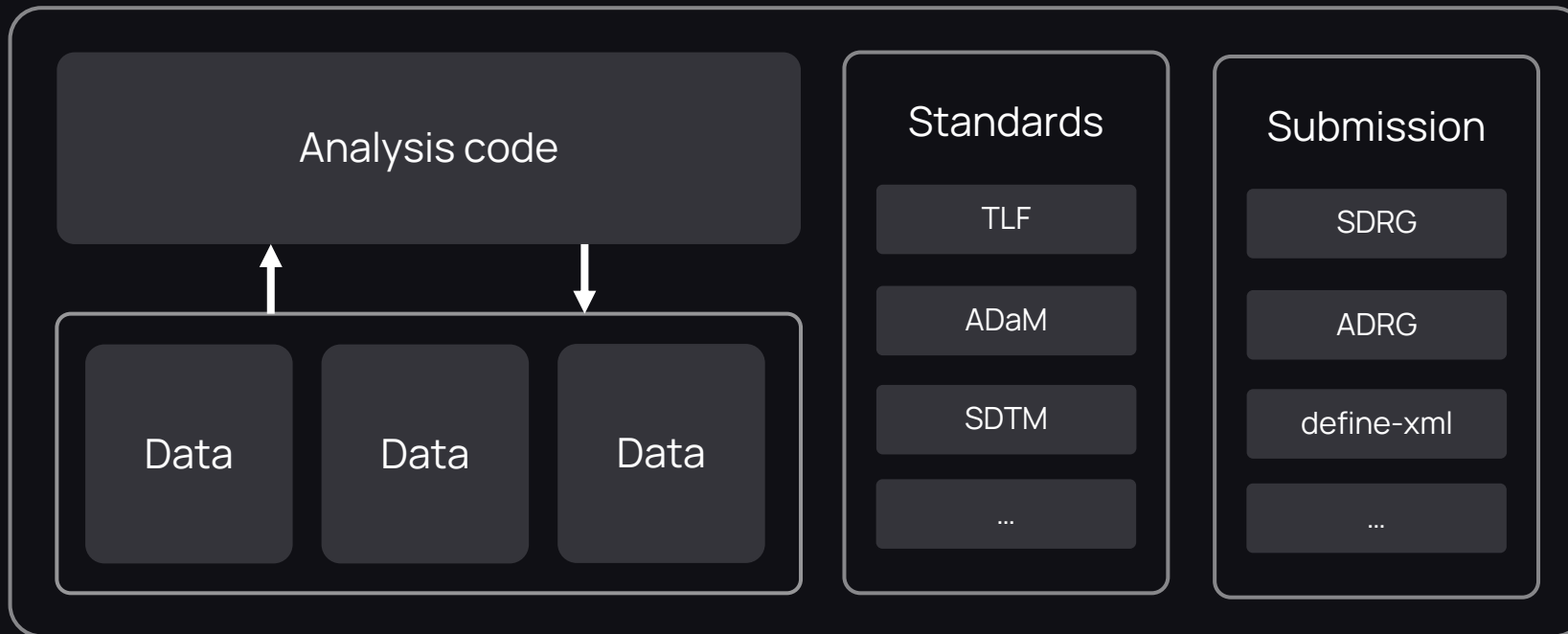


The challenge is to understand **data**
and **code**, and how they **interact**



Building Trust:
Validity, Traceability, Reproducibility

Once data are collected: what is a **clinical trial**?



Most derivations in a study are standardized

Derivations

Common derivations

...

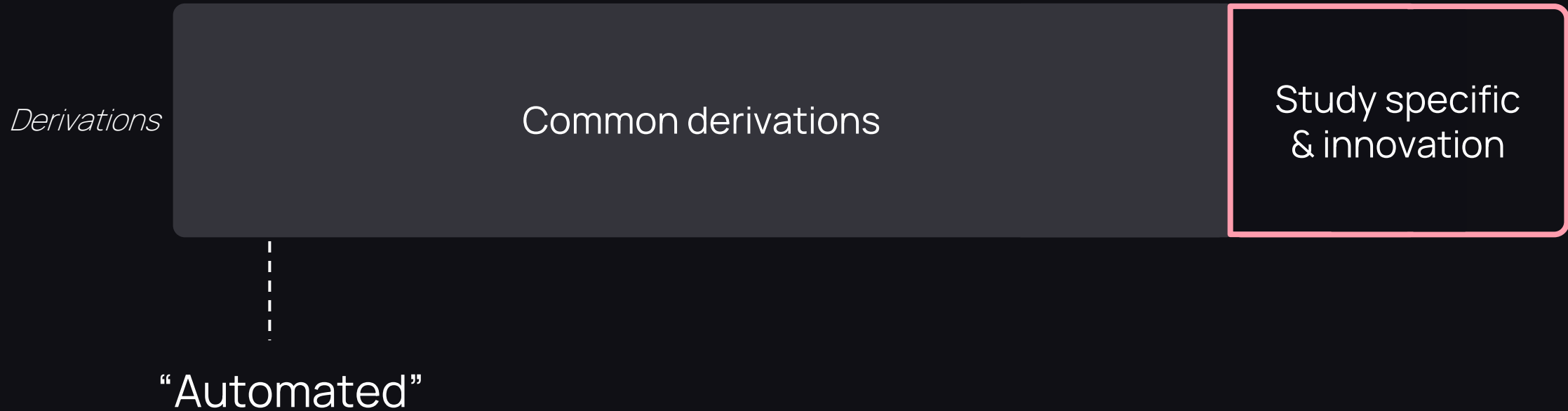
Most derivations in a study are standardized

Derivations

Common derivations

Study specific
& innovation

Most derivations in a study are standardized



Most derivations in a study are standardized

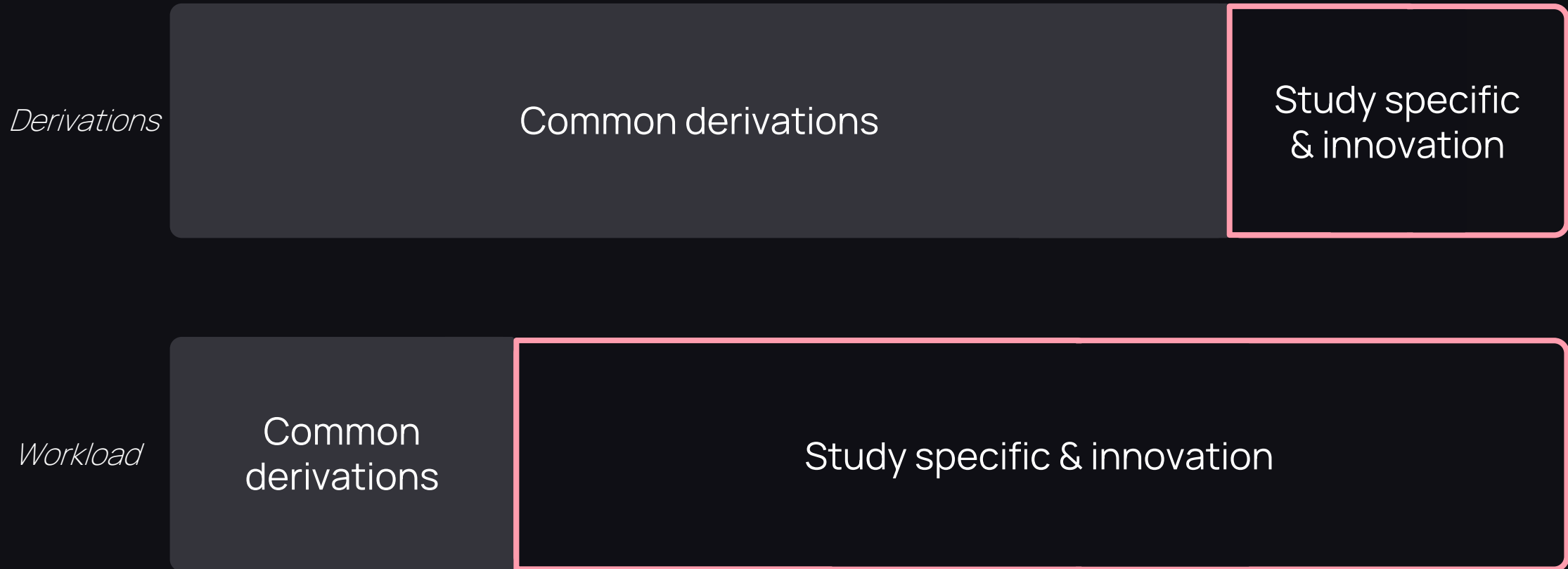
Derivations

Common derivations

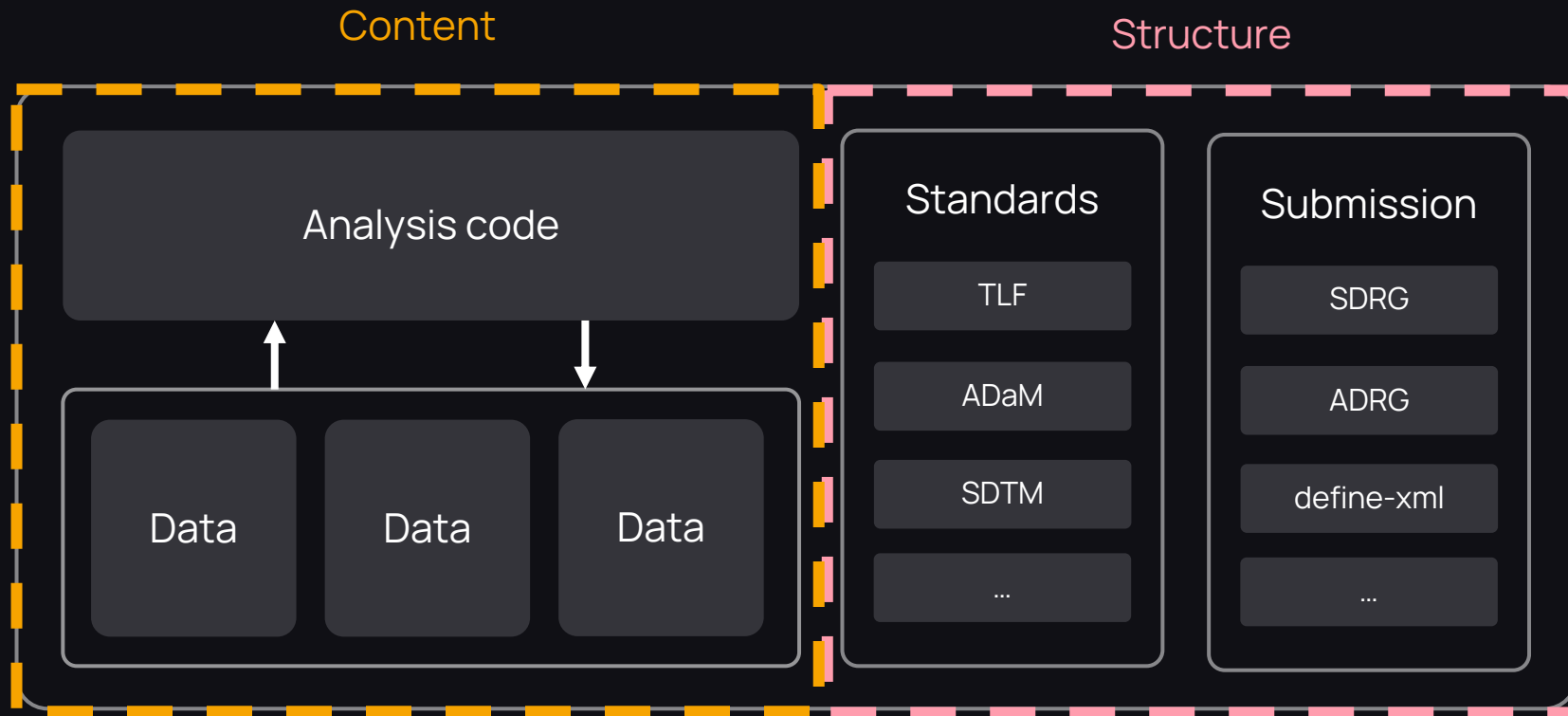
Study specific
& innovation

Not covered by standards, no traceability

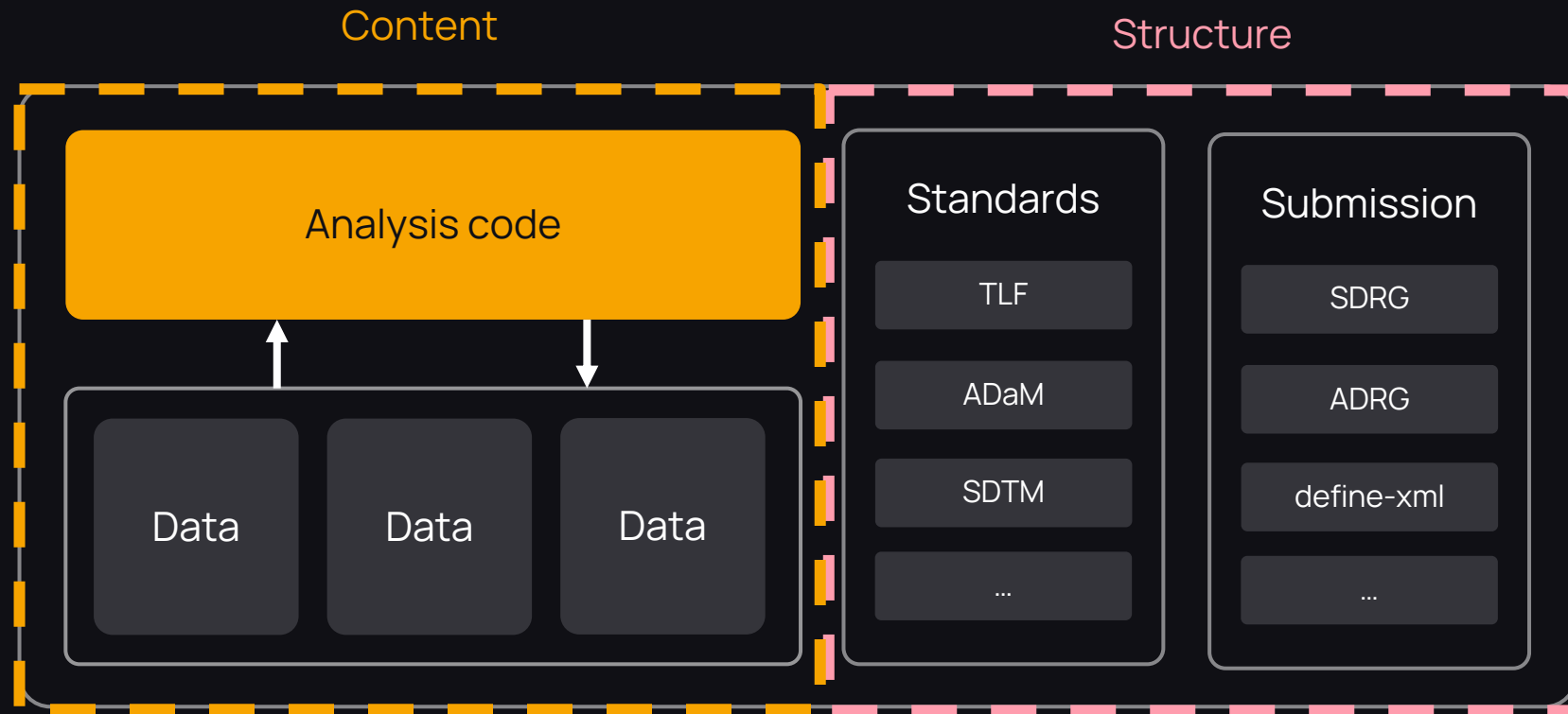
Most derivations in a study are standardized



What is a clinical trial?



The **source of truth** of **content** of a trial analysis is its **code**



What matters is the **content** of the study, not the **structure**

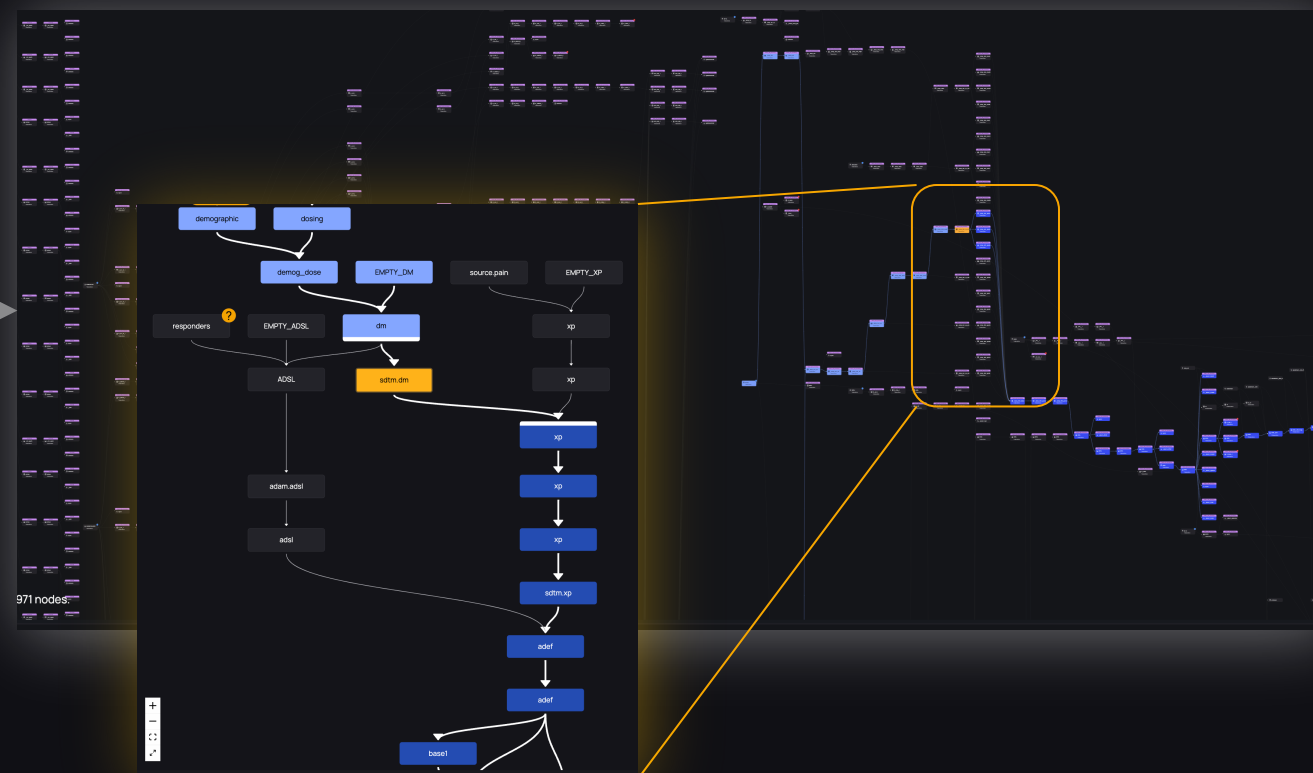
Non-standard, study-specific, scientific innovation

Code traceability

Tracing the content interactions between data, code and results

Independent of standards and languages

All you need are logs...



Q Search variables...

ADaM

adsl.arm

adsl.trt01p

adsl.trt01pn

adef.trtpn

SDTM

dm.arm

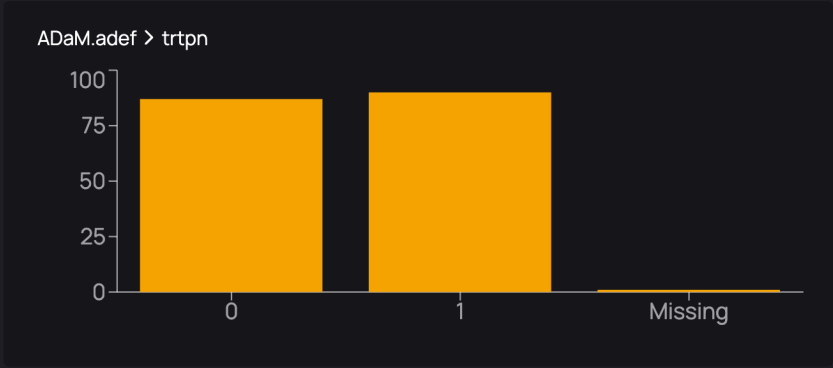
SOURCE

demographic.trt

Variable

Full Derivation

ADaM.adef > trtpn



trtpn	Count
0	90
1	90
Missing	2

(Macro-Resolved)

```
62 * tlf-responder-rate.sas;
63 *...;
64 run;
65
66 * adef.sas;
67 data adef;
68   retain studyid usubjid siteid country randdt age agegr1n a
69   set EMPTY_ADEF adef;
70 run;
71
72 * tlf-responder-rate.sas;
73 proc freq data=adef;
74   *...;
75   tables trtpn * crit1fl / chisq;
76   *...;
```

%make_empty_dataset

EMPTY_DM

0 observations

arm

dm

60 observations

arm

source.demographic

60 observations

trt

%make_empty_dataset

EMPTY_ADSL

0 observations

arm

trt01p

trt01pn

ADSL

60 observations

arm

trt01p

trt01pn

adef

177 observations

trtpn

%make_empty_dataset

EMPTY_ADEF

0 observations

trtpn

adef

177 observations

trtpn

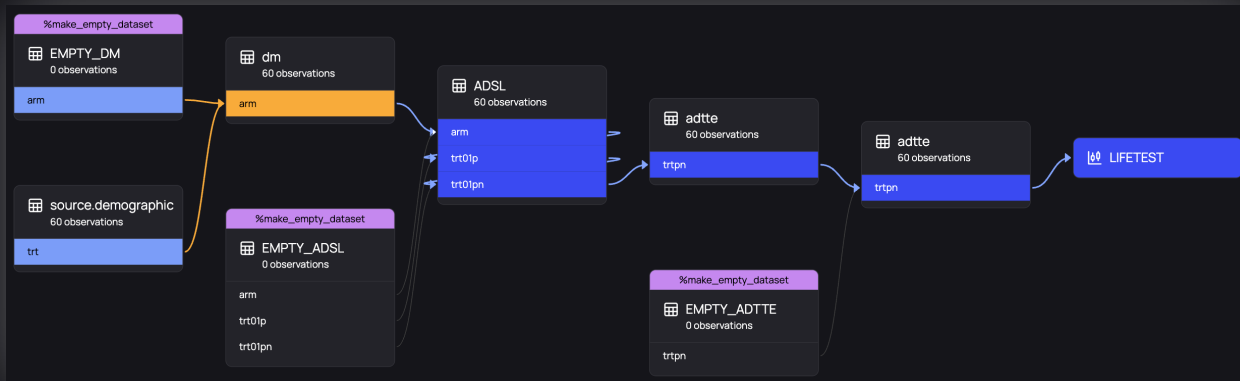
FREQ

GLM

10 nodes

Traceability-enabled AI

Creating an inclusive and deterministic environment



 Study Chatbot



Maria Shaw, Statistical Analyst 2 May 2024

How is the treatment emergent time frame defined?

  ads.keyds



Verisian AI 2 May 2024

In this code, trtemfl variable is used to determine the treatment emergence status of an event.

If the start date (astdt) is greater than or equal to the treatment start date (trtsdt), the event is considered treatment emergent and trtemfl is set to 'Y'.

If the start date is missing (astdt = .z) and the end date is either missing or on or after the treatment start date (not(.z < aendt < trtsdt)), the event is assumed to be treatment emergent and trtemfl is set to 'Y' as well.

An event is treated as treatment emergent if it started on or after the treatment start date.



 Ask a question...

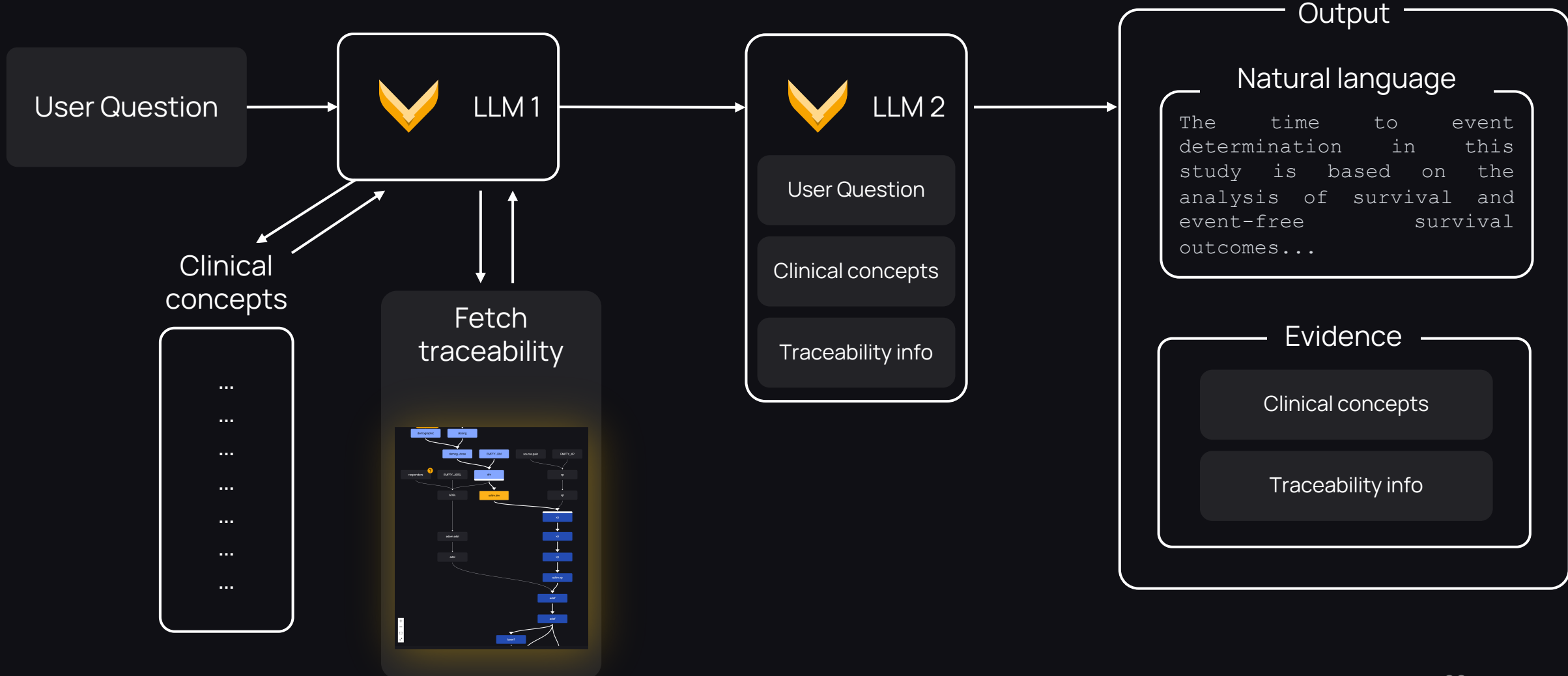
 Provide a context... 

Given any **natural language** prompt...

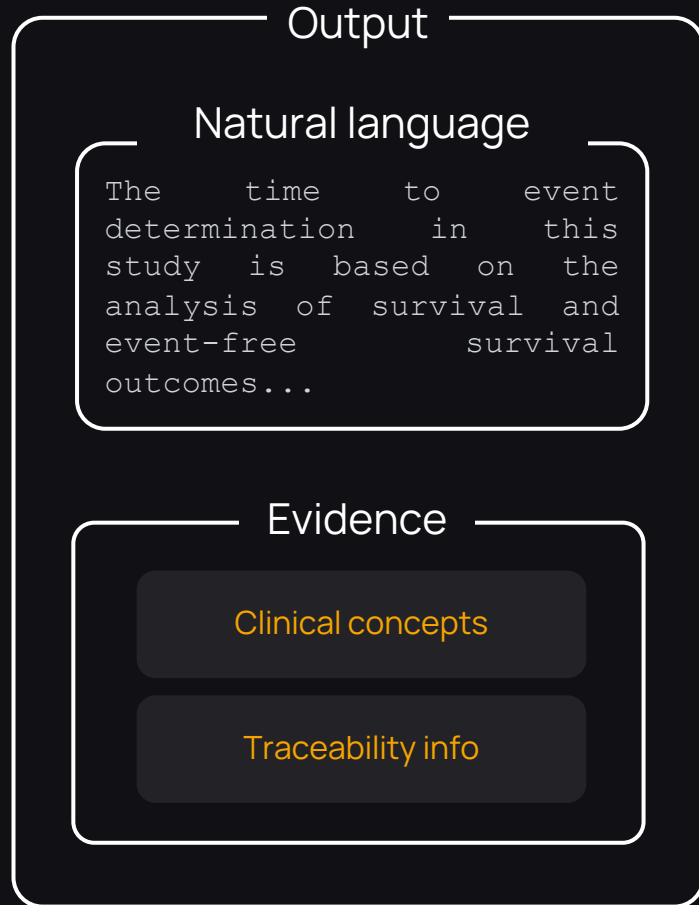
User Question

How is time to event
determined in this study?

...and translate into **natural language output**



...always provide **evidence!**



Deterministic

Translation of Code
Traceability

Evidence

Always provided to
review AI outputs

Study specific

Retrieving directly
from a study

Traceability-enabled AI

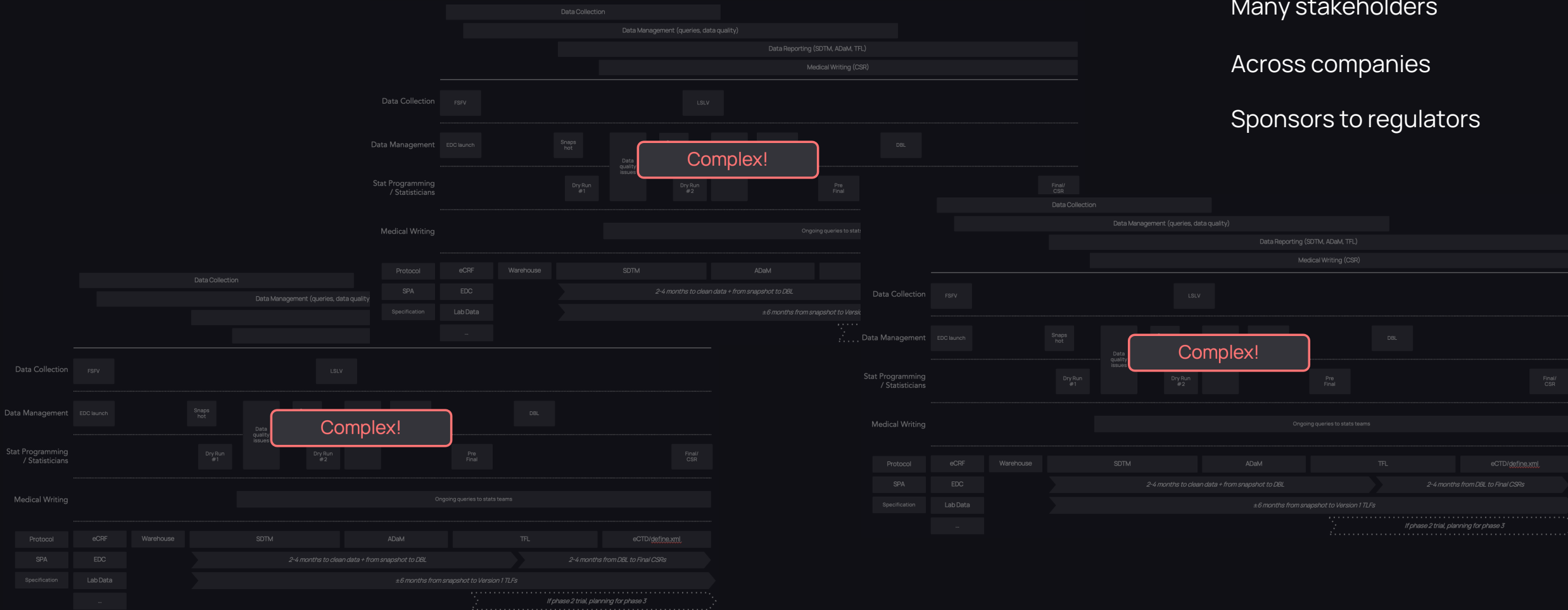
Reproducible, evidence based

Knowledge transfer in an integrated industry

Many stakeholders

Across companies

Sponsors to regulators





Tracer

Filters

(x) sdtm.dm.age X

Selection

(x) dosing.firstdose X



Graph Resolved Original Log

(Macro-Resolved)

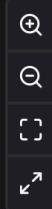
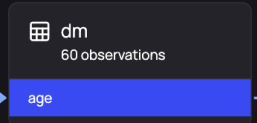
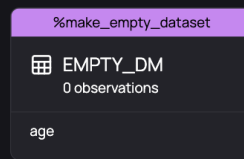
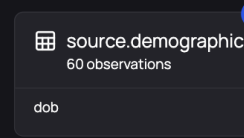
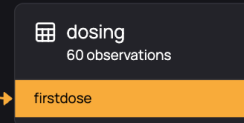


```
27     format startdt enddt mmddyy10.;
28     *...;
29 run;
30
31 * DM.sas;
32 data EMPTY_DM;
33     *...;
34     attrib AGE label="Age" length=8. ;
35     retain AGE . ;
36     *...;
37 run;
38
39 * DM.sas;
40 data dosing;
41     *...;
42     by subject;
43     *...;
44     retain firstdose lastdose;
45     *...;
46     if first.subject then do;
47         firstdose=.;
48         *...;
49         firstdose=min(firstdose,startdt,enddt);
50         *...;
51 run;
52
53 * DM.sas;
54 data dm;
55     *...;
56     age=floor ((intck('month',dob,firstdose) - (day(firstdose) < day(dob)
57     if age ne . then ageu='YEARS';
```



Graph Overview Graph

17 nodes



Variable Datasets

Name	Type
firstdose	unknown
Inferred Values	<unknown>

```

1 * source.demographic;
2 data source.demographic;
3   . label trt="Treatment";
4   input trt 5;
5   *...;
6 run;

7
8 * DM.sas;
9 data dm;
10  *...;
11  arm=put(trt, arm_demographic_trt.);
12  *...;
13 run;

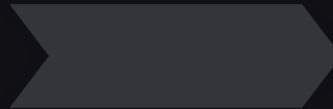
14
15 * adsl.sas;
16 data ADSL;
17  *...;
18  trt01p=ARM;
19  *...;
20  trt01pn=input(put(trt01p, $trt01pn.), best.);
21  *...;
22 run;

23
24 * adef.sas;
25 data adef;
26  * ...;
27  rename trt01pn=trtpn;
28  *...;
29 run;

30
31 proc freq data=adef;
32   by crit1 avisitn;
33   where ittfl='Y';
34   tables trtpn * crit1fl / chisq;
35   title
36     "Fishers Exact Test on Responder Rates Between Treatment Groups by Vi
37 run;
38

```

From 700+ lines to 37 lines



```

1 * demographic.sas;
2 data source.demographic;
3   . label trt="Treatment";
4   input trt 5;
5   *...;
6 run;

7
8 * DM.sas;
9 data dm;
10  *...;
11  arm=put(trt, arm_demographic_trt.);
12  *...;
13 run;

14
15 * adsl.sas;
16 data ADSL;
17  *...;
18  trt01p=ARM;
19  *...;
20  trt01pn=input(put(trt01p, $trt01pn.), best.);
21  *...;
22 run;

23
24 * adef.sas;
25 data adef;
26  * ...;
27  rename trt01pn=trtpn;
28  *...;
29 run;

30
31 proc freq data=adef;
32   by crit1 avisitn;
33   where ittfl='Y';
34   tables trtpn * crit1fl / chisq;
35   title
36     "Fishers Exact Test on Responder Rates Between Treatment Groups by Vi
37 run;
38

```

Using these techniques can
fundamentally transform
clinical statistical programming for submissions

Reproducible, evidence based

Build, Validate, and Submit Clinical Trial Analyses

Based on Traceability. Empowered by AI.

