MRC Clinical Trials Unit

UCL

Steps in using health systems data as outcome data in clinical trials

Sharon Love and Macey Murray

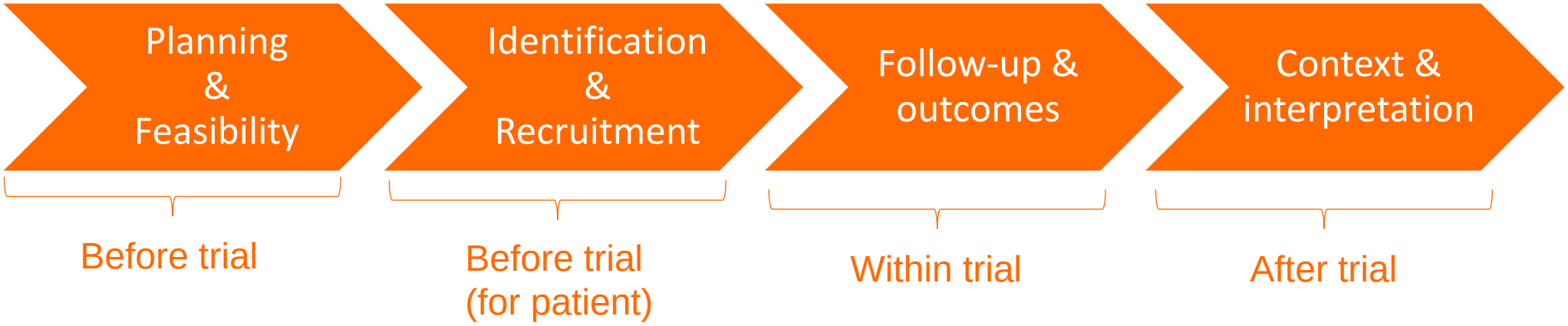
Smarter Studies
Global Impact
Better Health

PSI conference

11Jun2025

Increasing use of health systems data (HSD)

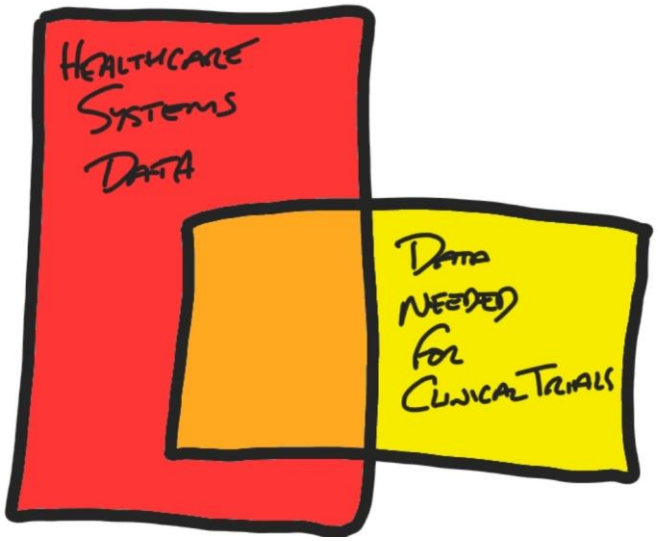
- RCTs = Gold standard for evidence-based medicine
- Using health systems data is one key stepping stones to streamline way to gain highest quality evidence

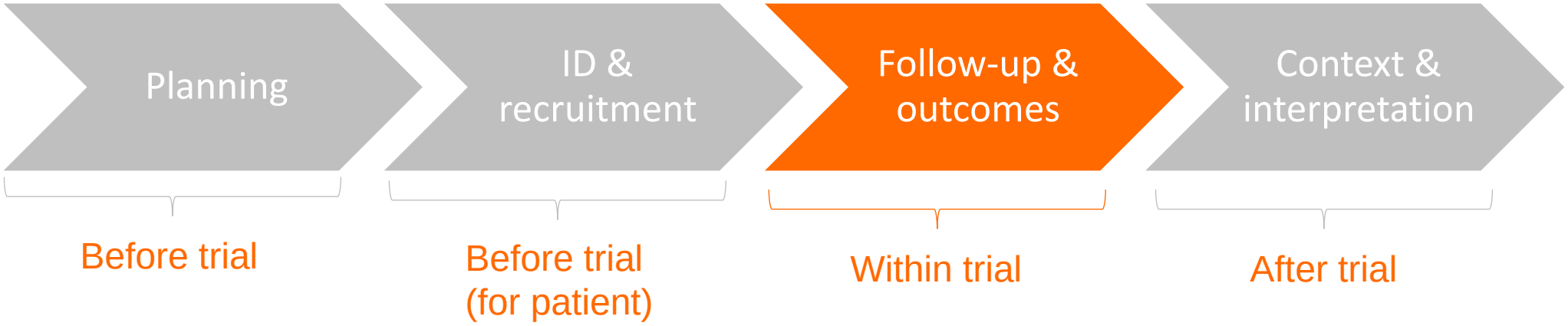


Timely & appropriate use of reliable, high-quality health systems data should transform how many clinical trials are done

Increasing use of health systems data (HSD)

- RCTs = Gold standard for evidence-based medicine
- Using health systems data is one key stepping stones to streamline way to gain highest quality evidence
- Considerable overlap between:
  → Data already collected in health systems
  → Data needed for clinical trials

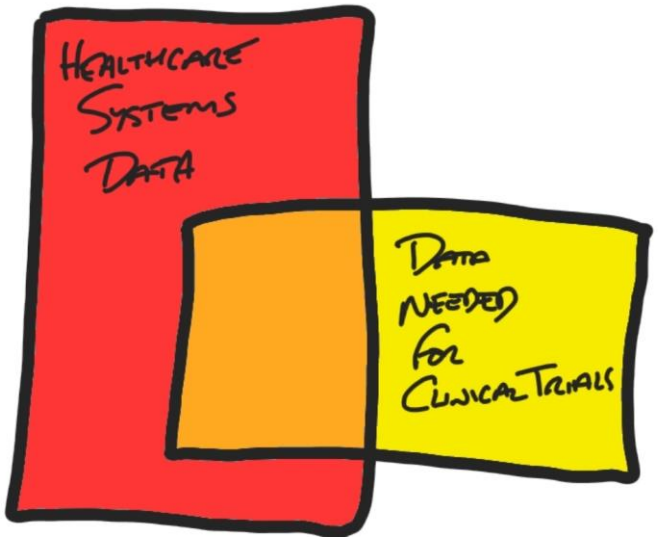


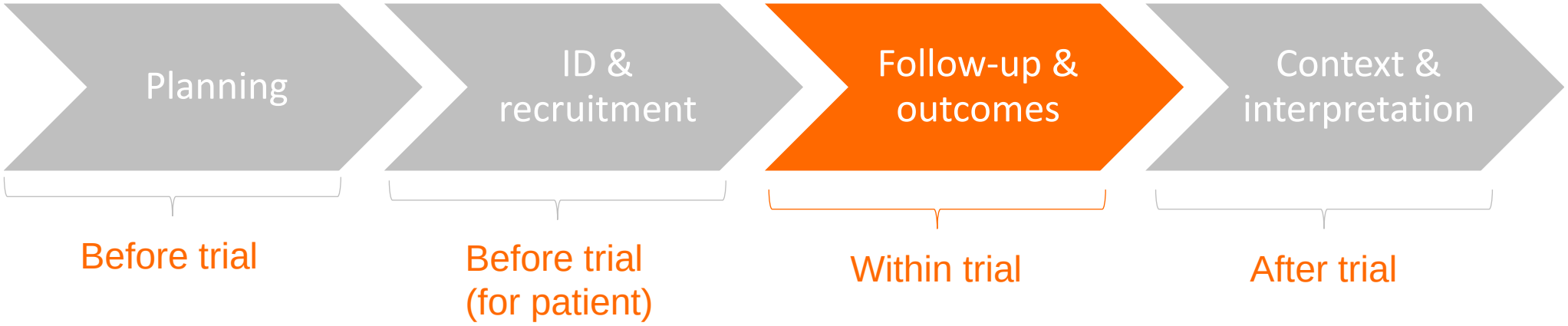


Timely & appropriate use of reliable, high-quality health systems data should transform how many clinical trials are done

Increasing use of health systems data (HSD)

- Reduce duplication of effort
- Reduce transcription errors
- Improve other aspects of quality (particularly completeness)
- Avoid missing important events (inc. Adverse Events)
- Allow team focus on trial data & systems that need attention
- Reduce trials costs (→ allowing more research)





Timely & appropriate use of reliable, high-quality health systems data should transform how many clinical trials are done

Challenges for HSD in clinical trials

- Knowledge about data & costs
- Access timelines & timeliness
- Integrity & provenance (data lineage)
- Utility (good enough to replace CRFs?)
- Analysis environment & federation
- Retention & archiving
- Onward data sharing

Challenges for HSD in clinical trials

- Knowledge about data & costs
- Access timelines & timeliness
- Integrity & provenance (data lineage)
- Utility (good enough to replace CRFs?)
- Analysis environment & federation
- Retention & archiving
- Onward data sharing

MRC Clinical Trials Unit

UCL

Provenance

shown by detailing the origins of the data, the processes and the methods by which it is produced

Integrity

shown by demonstrating the extent to which the data are complete consistent, accurate and reliable throughout the data lifecycle

Based on

- ALCOA+ principles (Accurate, Legible, Contemporaneous, Original, Attributable, complete, consistent, enduring, available when needed
- Clinical Data Interchange Standards Consortium (CDISC) eSource standard

Three stages

- collection and transfer of data
- centralised processing and curation to form the validated dataset
- linkage and extraction for trialists and the sponsor.

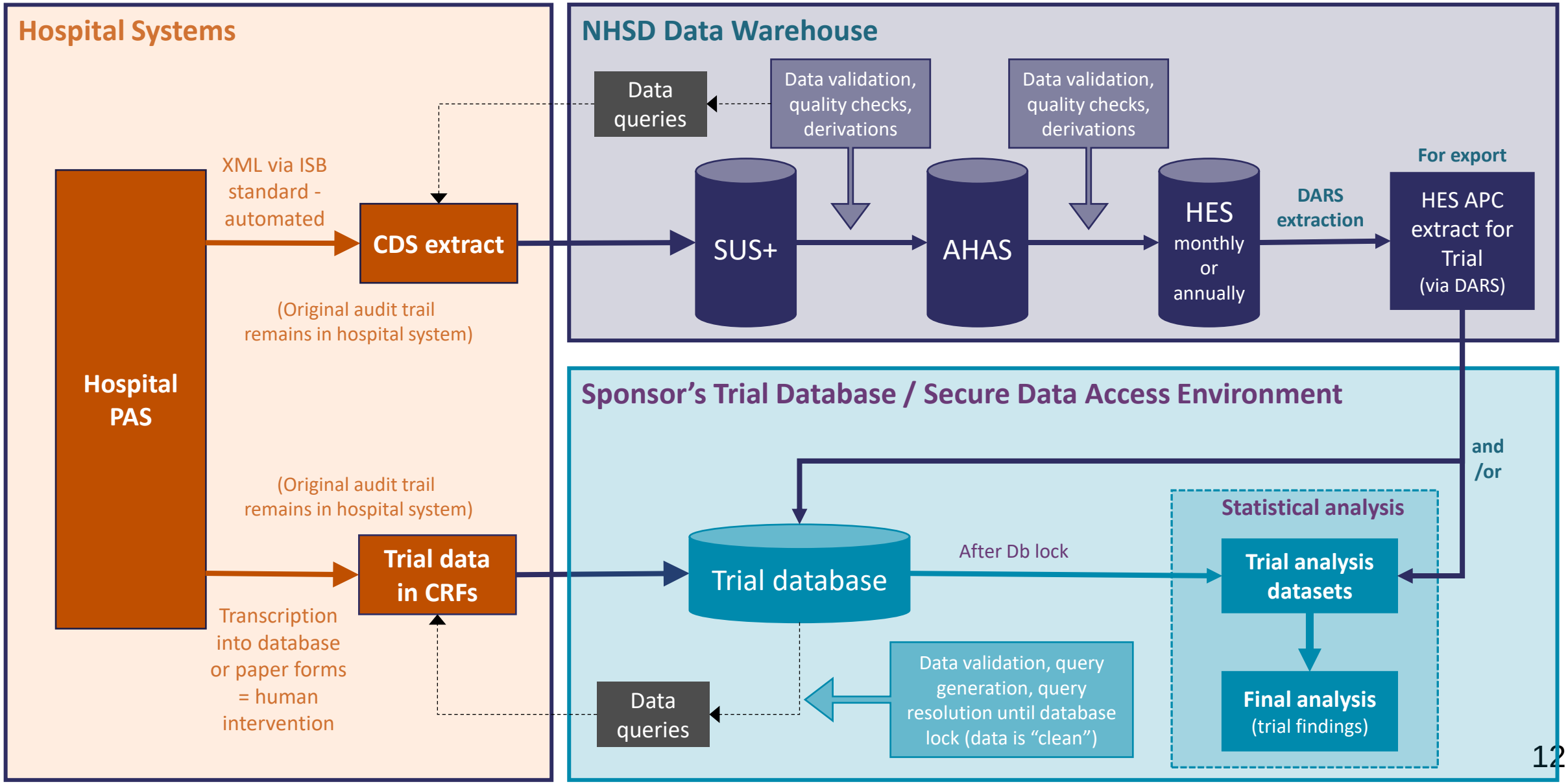
Sponsor needs to demonstrate

Trial sponsors need to demonstrate to regulatory authorities that all data, including healthcare systems data, are integral, reliable, and complete.

~

The data can be considered as equivalent to high-quality transcribed versions of the original source data.

Example – NHS England



Published work

Manual assessment of two UK data sets

Use of NHS Digital datasets as trial data in the UK: a position paper.

Macey L Murray et al

Zenodo. 2022. DOI:10.5281/zenodo.6047154.

How to automate

Demonstrating the data integrity of routinely collected healthcare systems data for clinical trials (DEDICaTe): A proof-of-concept study

Macey L Murray et al

DOI: 10.1177/14604582241276969

Utility

shown by looking for completeness and agreement between trial and health systems data, followed by showing that the trial specific data and health systems data give the same trial result

MRC Clinical Trials Unit

UCL

Utility

Broad question

- Agreement of fact of data item and, if so, agreement on its timing and definition

Method

- SWAT using Kappa if binary

Barriers

- Neither trial or health system data is the obvious gold standard
- Need trial and health system data in same location

Unexpected benefit

- Allows evaluation and exploration of discrepant events.

Utility of death data

SWAT 125: Comparison of trial-collected and routinely-collected death data

Available from:
https://www.qub.ac.uk/sites/TheNorthernIrelandNetworkforTrialsMethodologyResearch/FileStore/Filetoupload,976743,en.pdf

Preliminary results: Fact of Death

Pt level comparison

Deaths reported	STAMPEDE trial specific data		
HSD	Alive	Died	Total
Alive	890 (98.6%)	167 (5.3%)	1057 (26.2%)
Died	13 (1.4%)	2960 (94.7%)	2973 (73.8%)
Total	903	3127	4030

Agreement in (890+2960)/4030 (95.6%) in the number of deaths

$\kappa = 0.88$

MRC Clinical Trials Unit

UCL

HSD – Health Systems Data

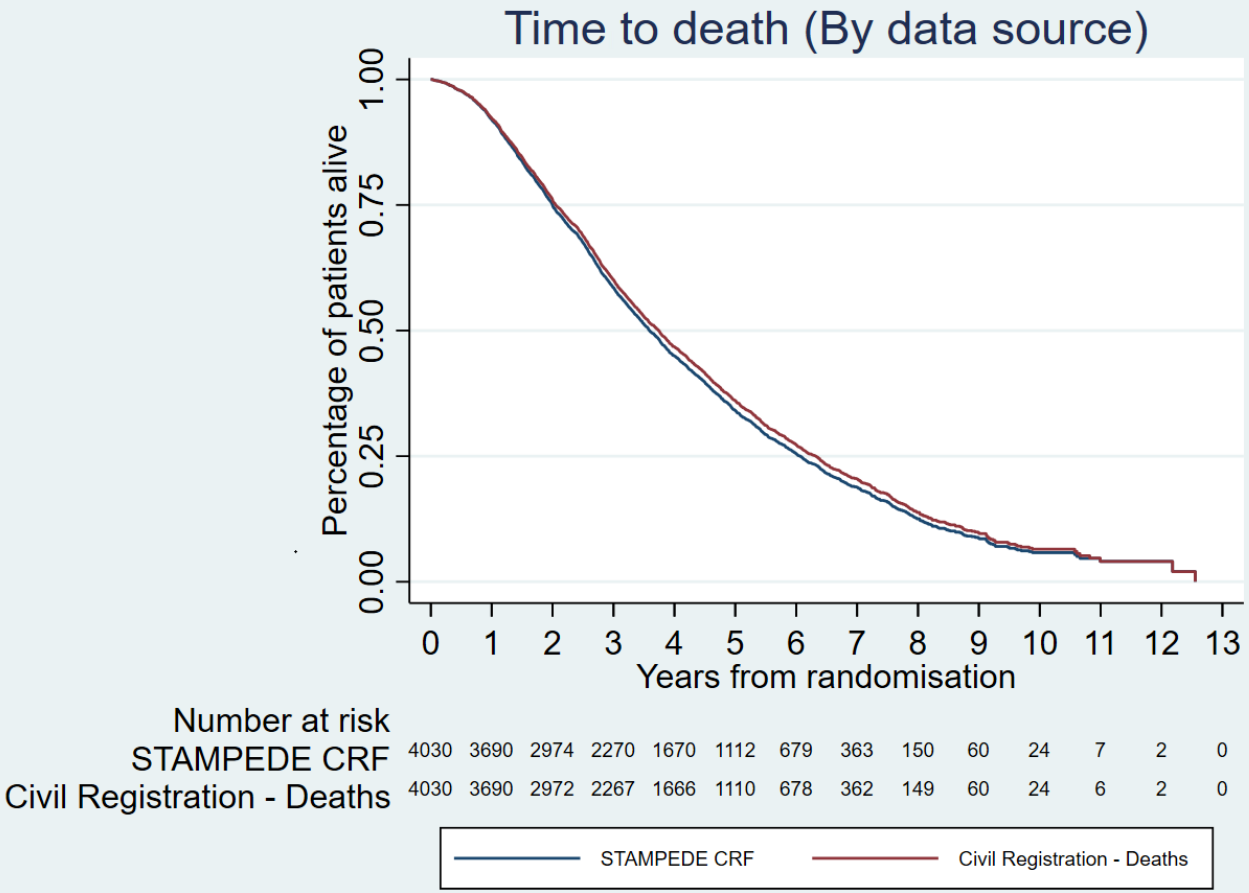
Preliminary results: Agreement in death date reporting

Pt level comparison

Number (%) date agreement between HSD and CRF	Difference in dates	Frequency
	Exact agreement	2781 (94.0%)
HSD date < CRF date	1 – 7 days	159 (5.4%)
	8 – 14 days	12 (0.4%)
	15 – 30 days	4 (0.1%)
	31 days or more	4 (0.1%)
	Total	2960

Preliminary results: Time to death

Summary measure level comparison



Median time to death in both sources were similar

STAMPEDE CRF

3.6 years (95% CI 2.00 – 6.06)

HSD

3.7 years (95% CI: 2.05 – 6.32)

MRC Clinical Trials Unit

UCL

Health System Data is from Civil Registration - Deaths

Example summary

1. Comparison of death data between CRF and HSD relatively straightforward.
2. Some differences in patient level comparison, however similar conclusion at the summary measure level.
3. Future trial design can consider replacing death data collection though the HSD.

MRC Clinical Trials Unit

UCL

Useful references/training

Getting our ducks in a row

https://doi.org/10.1016/j.cct.2024.107514

Data Utility Considerations for Clinical Trials

https://hdruklearn.org/

Summary

Health Systems Data is useful for outcomes in clinical trials

Checks are required

Methodology for these checks is available

MRC Clinical Trials Unit

UCL