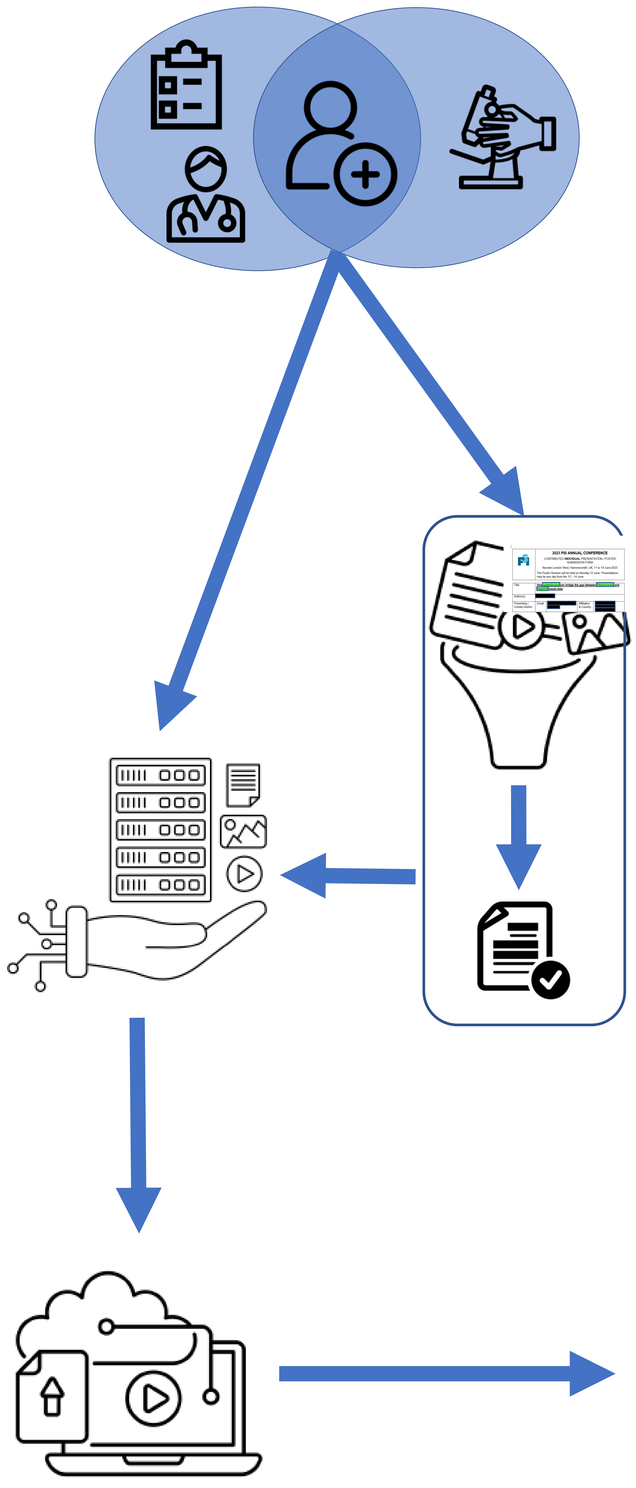

How statisticians can bridge the gap between Clinical trials and Registry based data

Andrew Mills, Catherine McHugh, Jamie Inshaw, Alessandra Bisquera, Amy McCorry & Sheila McCartan
Exploristics

1 Objectives

- Collect data directly from **participants** as well as **clinical sites**
- **Extract** and **de-identify** data from **unstructured** sources
- Build data collection techniques that are **generalised** to suit a **non-interventional** protocol.
- Create a **standardised Registry database** for ease of use by the wider scientific community

2 Approach

- Utilise a variety of data sources collected within the registry (**structured** and **unstructured**)
  - eCRF
  - Patient centric data capture tools
  - Physician letters
  - PDF lab reports, etc.
- Build in **automated processes** for data collection, extraction and validation
  - Natural Language Processing algorithms
  - Machine Learning models
- **Augmenting data** with **external sources**.
  - Clinical Trials registry (ClinicalTrials.gov)

3 Methods

- Mirror data collection captured from **clinical sites** and directly from **patients**
  - Use the site collected data to **validate** the process of collecting data from patients
- Incorporate efficient **data observation processes** to assess validity of data
  - Risk Based and Centralised Monitoring
  - Automated data checks
  - Using simulation to determine appropriate thresholds for impact of missing data
- Validate baseline characteristics by comparing inclusion / exclusion criteria of clinical trials that Registry participants are involved in using **National Clinical Trial numbers**
- Performing manual data **extraction** and **de-identification** from unstructured sources to use as data for test, training and validation of an automated process
- Using fit for purpose **dataset standards** such as CDISC

4 Impact

- Statisticians can
  - **Influence** and **impact** on the strategic planning
  - Highlight appropriate statistical techniques or methodologies to provide (and validate) **alternative processes** of obtaining information

Transforming independently captured information into a versatile ecosystem.

References:

Adamson et al, *Approach to Machine Learning for Extraction of Real-World Data Variables from Electronic Health Records* (2023)

Acknowledgements:

Kimberley Hacquoil & Sam Miller **Exploristics**

exploristics

unlocking the value in data