MRC-NIHR Trials Methodology Research Partnership: Webinar recording

# The PRIMORANT study:
# When can routinely collected data replace bespoke data collection?

*Presented, on behalf of Health Data Research UK, by:*

Amanda Farrin (University of Leeds), Alice-Maria Toader (University of Liverpool)
and Paula Williamson (University of Liverpool)

**15 June 2023**

The slides are available below.

For any queries, please contact uktmn@nottingham.ac.uk

https://youtu.be/nNwmLnCulYk

# The PRIMORANT study

Addressing high priority methodological questions for routinely-collected healthcare data and trials

## When can routinely collected data replace bespoke data collection for trial outcomes?

**Amanda Farrin, Alice-Maria Toader, Paula Williamson**

# Acknowledgements: the PRIMORANT team

HDR-UK North, Universities of Liverpool & Leeds:

**Paula Williamson, Amanda Farrin, Alice-Maria Toader,** Munir Pirmohamed, Andrew Clegg

CTR, Cardiff University: Fiona Lugg-Widger, Mike Robling, Julia Townson, Rob Trubey

MRC CTU at UCL + BHF Data Science Centre: Matthew Sydes

MRC CTU at UCL + NHS Digital: Macey Murray

University of Oxford CTSU: Marion Mafham

HDR UK Training Team: Sarah Cadman, Timothy Frayling, Rosie Wakeham

In collaboration with the Trials Methodology Research Partnership (TMRP) Health Informatics Working Group Routine Data Topic Group

Funded by HDR-UK Directors' Discretionary fund

# Polls 1

Please tell us about your role in trials

Cardiff University
Prifysgol Caerdydd
Centre for Trials Research
Canolfan Ymchwil Treialon

COMORANT UK
A CONSENSUS ON THE METHODOLOGICAL OPPORTUNITIES FOR ROUTINE DATA AND TRIALS

# COMORANT-UK: The prioritised seven questions to address

**The COMORANT-UK study aimed to systematically identify, with key stakeholders across the UK, the ongoing challenges related to trials that seek to use routinely-collected data.**

This 3-step Delphi method consisted of two rounds of anonymous web-based surveys, and a virtual consensus meeting.

Stakeholders included trialists, health relevant data infrastructures (i.e. HDR UK), funders of trials, regulators (HRA, MHRA), data providers and the public.

These prioritised seven questions address both evidence gaps (requiring further methodological research) and implementation gaps (requiring training and/or service re-organisation).

## Trial Design

**Data collection method**

When is it more efficient, considering trial design, costs, time and environment, to use routinely collected datasets compared to bespoke data collection?

## Trial Design

**Outcome selection**

How should the trials community decide when routinely collected data for outcomes is of sufficient quality and utility to replace bespoke data collection?

## Patient and Public Involvement

**Communication**

What are the best methods to communicate and build trust with participants (and the public) about how their routinely collected data will be used?

## Trial Set-up

**Regulatory Approvals**

How can approvals at trial set-up be streamlined across regulatory and data provider applications?

## Trial Open

**Data access and receipt**

How can routinely collected data flow (approval through to data provision) from all providers of data be expedited for analysis?

## Trial Data

**Quality**

What causes inconsistencies in routinely collected data across sources and how can these be identified, managed and reconciled for key trial outcomes (e.g. fact and date of death)?

## Trial Data

**Analysis**

Why are data missing in routinely collected datasets (person and individual data fields) and how should this inform methods for managing missing data?


**Collaborators**

University of Oxford · UCL · University of Leeds · UKCRC Registered Clinical Trials Units

**Funders**

MRC Medical Research Council · NIHR National Institute for Health Research · Trials Methodology TMRP Research Partnership

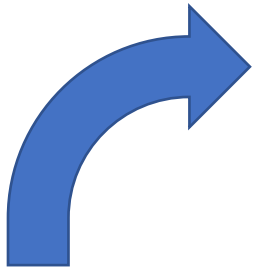
# PRIMORANT – 2 workstreams

## Trial Design

**Outcome selection**

How should the trials community decide when routinely collected data for outcomes is of sufficient quality and utility to replace bespoke data collection?

HDR UK North
Establishing
best practice



Cardiff
Training

## Patient and Public Involvement

**Communication**

What are the best methods to communicate and build trust with participants (and the public) about how their routinely collected data will be used?

# Polls 2

Have you used routine data for outcomes in a trial?

# Workstream : CONSORT-Routine: Outcomes

## Item 6a (modified)

Completely defined pre-specified primary and secondary outcome measures, including **how and when they were ascertained and the cohort or routinely collected database(s)** used to ascertain each outcome.

## Item ROUTINE-5 (new)

Information on how to access **the list of codes and algorithms used to define or derive the outcomes** from the cohort or routinely collected database(s) used to conduct the trial,

and

Information on **accuracy and completeness of outcome variables, and methods used to validate accuracy and completeness** (eg, monitoring, adjudication), if applicable.

# Workstream : SPIRIT-Routine: Outcomes

McCarthy 2022 HRB Open Research
doi: 10.12688/hrbopenres.13314.1

## Item 12 (modified)

Primary; secondary; and other outcomes including how and when they were ascertained and the cohort or routinely collected database(s) used to ascertain each outcome. If applicable; include the specific measurement variable (e.g.; systolic blood pressure); analysis metric (e.g.; change from baseline; final value; time to event); method of aggregation (e.g.; median; proportion); and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended

## New item

Information on how to access the list of codes and algorithms that will be used to define or derive the outcomes from the cohort or routinely collected database(s) that will be used to conduct the trial; information on accuracy and completeness of outcome variables; and methods that will be used to validate accuracy and completeness (e.g.; monitoring adjudication); if applicable.

# Terminology

Routinely collected data $\longrightarrow$ Healthcare systems data (HSD)

# Workstream 1: Progress to date

**Virtual workshop**: September 2022

- 26 attendees
- Identified areas / examples of best practice
- Discussed areas requiring agreement on guidance to trialists using HSD for outcomes
- Informed stakeholder consultation topics

**Consultation survey of key stakeholders**: Dec 2022 – Jan 2023

- 85 responses from CTU staff & trialists, methodologists, clinicians, funding panels, data providers

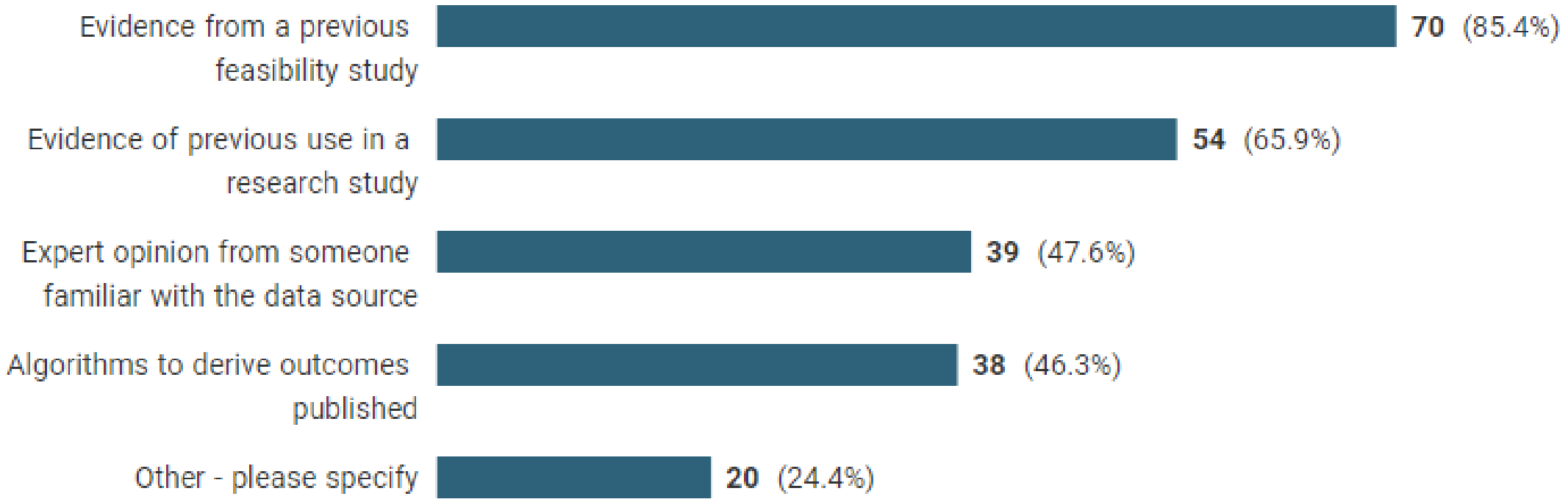
**In-person discussion workshop**: March 2023

- 35 attendees: statisticians, data managers, health economist, funding panel members, clinician, data provider
- Identify areas of uncertainty; challenges & solutions; examples of best practice
- Presented survey results & case studies from 7 trials
- Breakout groups discussed
  - Areas for guidance to trialists using HSD for trial outcomes
  - Draft wording for guidance (or points to consider)
  - Test usability against current experience

# Results from consultation

# 1. Validity of outcome data

What evidence would you consider sufficient to establish the reliability/validity of HSD for trial outcomes?




*Multi answer: Percentage of respondents who selected each answer option (e.g. 100% would represent that all this question's respondents chose that option)*

# 2. Timeliness of data capture for target outcomes and frequency (actual and expected) of data receipt

Are there situations where you have discounted/would discount using HSD?

- When aware of the delay in getting the data
- Safety of treatment need careful monitoring
- Cost efficiency
- Insufficient data or quality of data
- Insufficient information about the data

## Would a common SOP for resolution of discrepancies for data providers be helpful?


## Would a common SOP for (all) data providers be at all feasible?


## Would a template/SOP/guidance for trialists be (any) more feasible?

# What do you think is key to include in such a common SOP?

- Resolution of discrepancies
- Agreeing response timeliness
- Data handling & algorithms
- Missing data
- Data quality assurance (including any data linkage)
- Terminology
- Considerations – questions to consider
- Error handling
- Adverse effects

# 3. Internal pilot to look at validity of routine data for trial outcomes

Are you aware of trials using HSD for outcomes where an internal pilot would be/would have been helpful?


What Stop/Go/Amend progression criteria might you consider in such an internal pilot?

- What data is available?
- Missing data
- Time to access the data
- Quality of data
- Successful linkage
- Bias

# 4. Lack of published detail about practical issues related to use of HSD for trial outcomes

Which practical/logistical aspects of using routine data would be most helpful to publish?

- Cost
- Time – including approximation for each step of the process
- Processing challenges and methodology
- Quality of data – including data validity and Missing data
- Case studies and FAQs
- Feasibility and validity studies

# 5. Decision-making

What decision criteria would you use to decide between using HSD for trial outcomes vs using more traditional ways of collecting data for trial outcomes vs HSD being supplemented by more traditional data collection methods?

- Cost and resources needed
- Outcomes of Interest
- Quality of data
- Time to access the data
- Feasibility
- Transparency of data
- Patient burden
- Linkage
- Difficulty of process
- Bias

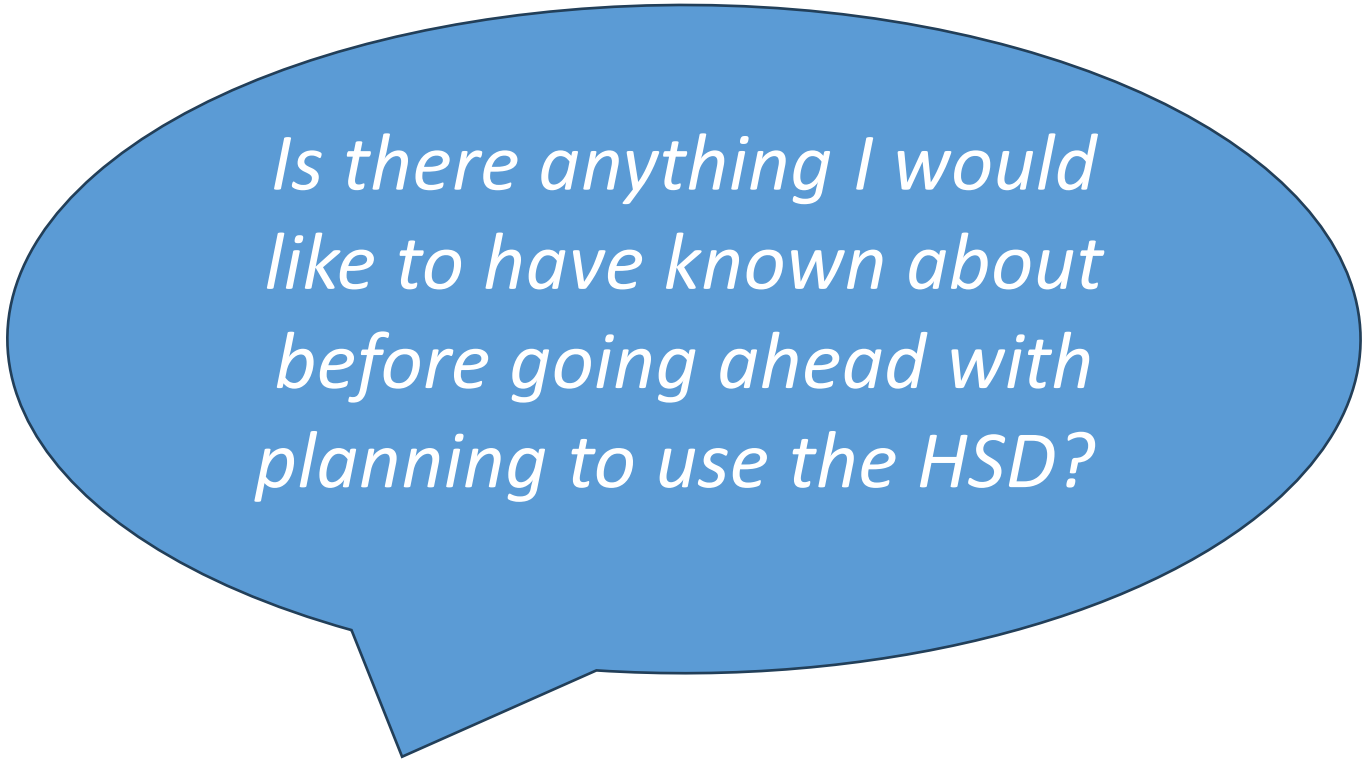
# Purpose of the workshop

How should the trials community decide when routinely collected data for outcomes is of sufficient quality and utility to replace bespoke data collection?

- Identify areas of uncertainty; challenges & solutions; examples of best practice
- Agree areas for guidance to trialists to be used before the decision to use HSD for collecting outcomes is finalised
- Develop draft wording for guidance (or points to consider)
- Test usability against current experience

# Reflection

For those with experience of using HSD for outcome data in trials, or those who have been thinking about it for a future trial:



# Is this list complete?

1. **Feasibility** (prior evidence, resources & cost)
2. **Data quality** assurance (including any data linkage)
3. **Time** (is there sufficient time to receive data? How often?)
4. **Internal pilot** considerations
5. **Terminology** (common understanding of definitions / principles / terms / measures )
6. **Algorithms** for deriving trial outcomes
7. **Data handling** (what happens to data between clinical data entry & researcher data receipt)
8. **Adverse effects** (when suitable for monitoring adverse outcomes?)
9. Process for handling **missing data**
10. Process for handling/reporting **errors in HSD**
11. **Raw data** or analysis ready data-sets?
12. **Data sharing considerations** (both DSAs & onward sharing)

# Poll 3

Is there something missing from this list that you wish you'd known or explored in advance of starting your trial using HSD for outcomes?

# Workshop results – issues to consider

(1) Terminology

(2) Feasibility

(3) Internal pilot

(4) Onward data sharing

(5) Data destruction and archiving

# Scope

- To be used before the decision to use HSD for collecting outcomes is finalised
- To help team to make an informed judgment based on an understanding of the suitability of HSD
- To help to identify uncertainties and build in mitigation

# (1) Terminology

- Understand terminology used (which may differ across different organisations)
- Be aware that terms used within data access applications will likely differ between providers

- seek examples from the provider if available

# (2) Feasibility

**Trial team**

Include data and disease specialists with experience of (i) completing data access forms and (ii) analysing the particular health data from the provider, those who:

- understand the data, its structure, and its quality
- understand how and when the data are collected at source
- have the skills to handle the data when it is provided
- undertake the statistical and health economic analysis

## Data: (a) Has it got what you need?

- Using the provider data dictionary where available, establish which outcomes are collected “routinely”
- Ascertain cost of data provision and timelines for data verification/release
- Discuss the process for data linkage if needed

- If time and resources permit, interrogate the dataset prior to the decision to use HSD

## Data: (a) Has it got what you need?

- If HSD covers only a subset of the outcomes, consider how the rest will be collected
- Registry-based trial - is registry team able and willing to adapt data collection to meet the trial's needs?
- In terms of using HSD for reporting safety data, consider the risk profile of the trial
- Public support – Existing evidence for support for accessing these data for research, or have issues arisen previously?

## (b) Data quality assurance

- Does provider have information regarding data provenance, integrity, and completeness?
- Timeliness of data held – is there a lag between site data collection and entry into the provider system?
- Understand how provider receives and processes the data, and how changes in processing and coding are handled and communicated
- Previous literature – what is already known about the validity and completeness of the outcome data

- Data quality – realistic to provide funder with accurate idea of data quality at application, or possible to build in approach to examine uncertainty during the trial?

## (c) Time

- Ask provider - how long from point of request and then from point of approval to supply a specified dataset?
- Determine if contract includes binding timelines; decide what is an acceptable delay
- Consider in relation to whether any interim analyses are planned or when using HSD for monitoring safety outcomes.

## (d) Algorithms for deriving outcomes

- Explore whether a validated algorithm for deriving outcomes from HSD exists
- If not, consider whether to include time to develop and test the proposed algorithm, within a utility comparison.

## (e) Considerations around missing data

- Important to appreciate if available data not collected at expected time, e.g. annual reviews may be delayed
- Discuss whether it may be possible to go back to sites to collect missing data
- Consider imputation from other variables, or other HSD datasets - contingency fund?

## (f) Consideration of potential reporting errors/discrepancies

- Discuss mechanism and opportunity for discrepancy resolution with provider
- Ask provider for their experience of possible solutions, e.g. rules of precedence, windows for ‘same dates’, impossible events
- Managing data queries – contingency fund?

## (g) Preparation of trial dataset

- Discuss with provider - raw data or analysis-ready dataset?

# (3) Internal pilot

- Is use of HSD more cost-effective than collecting outcome data traditionally?

- Stop/Go/Amend criteria (related to use of HSD)

- What data are available?
- Amount of missing data
- Time to access the data
- Accuracy of the data
- Success of linkage

- Can using both approaches be justified in terms of cost and complexity
- e.g. by providing added value for the health area more widely than the individual trial

# (4) Onward data sharing

- Important to consider who (e.g. trial oversight committees, trial team, industry partners, future meta-analysts) will need to see HSD, as raw or aggregated data
- Discuss funder's requirements for onward data sharing – can provider approve this
- Onward sharing may not be permissible or subsequent access may not be straightforward
- Ensure these issues are considered in the data sharing agreement/contract
- Explore legal, ethical and governance responsibilities in advance
- Consider implications for consent forms, allowing further use of data past the initial trial.

# (5) Data destruction and archiving

- Discuss any regulatory requirements for the archiving period with the data provider

# Benefits of guidance

- May help to build in mitigation, e.g. supplement with data directly from participants or sites
- Consideration of additional costs that could be incurred or unanticipated workarounds required

- changes in legislation
- delays in data release
- periodic renewal of data sharing agreements

- Inform discussion with funder - contingency fund, phased project plan, etc
- May highlight ways to work with providers to improve how trials designed and delivered

# What's next?

- Paper in draft
- Share guidance
  - HDR-UK
  - Data providers
  - Funders
  - CTU network
- Feedback welcome

Trials Methodology TMRP Research Partnership

@MRCNIHRTMRP

https://www.methodologyhubs.mrc.ac.uk/about/tmrp//

Health Research Board TMRN Trials Methodology Research Network

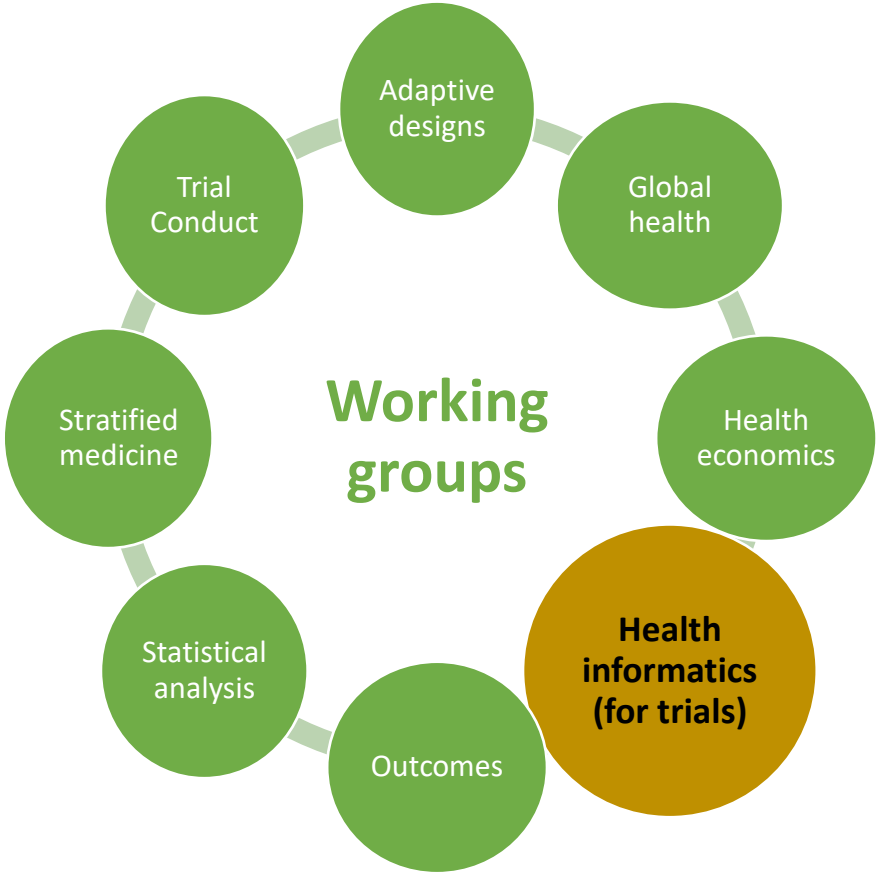
MRC Hubs for Trials Methodology Research

HDRUK Health Data Research UK

UKTMN Trial Managers' Network

UKCRC Registered Clinical Trials Units

THE GLOBAL HEALTH NETWORK
Enabling research by sharing knowledge




**HIWG: methodology to realise potential of digital technology to improve trial design, conduct and analysis**

PPIE

25 universities
with strong trials methodology expertise

**Routine data, participant data, data sharing, digital technology, digital health technologies**

**Trial Design**

PhD students in methodology

MRC Doctoral Training Programme