Challenges and disincentives for sharing and using data: A researcher perspective

Professor Matthew Sydes
MRC CTU at UCL
Institute of Clinical Trials and Methodology
London, UK

Sharing Clinical Trial Data: Challenges and a Way Forward: A Workshop
17-Nov-2019
Data sharing at ICTMC 2019

• ICTMC 2019 = conference on clinical trials methodology
  – 730 registered
  – 500 abstracts

• Only 4 (<1%) related to clinical trials data sharing
  – Applicant experiences of CSDR
  – Systematic review of “anonymisation methods”
  – Do data silos jeopardise IPD meta-analyses
  – CTU changing approach

Some Challenges

Using data in a valid way and usefulness of new knowledge generated from data sharing

Rationale for linking to trial data and linkage to routine data for sharing

Oversight processes necessary for access to shared data

Recognition of researchers involved in creating data
Some Challenges

Using data in a valid way and usefulness of new knowledge generated from data sharing

Rationale for linking to trial data and linkage to routine data for sharing

Oversight processes necessary for access to shared data

Recognition of researchers involved in creating data
Using data in valid way & utility of new knowledge

• Need: maximize highest quality science to have greatest societal benefit
  – Much from clinical trials
  – Much from associated research

• Conscious of opportunity costs

• Effort to prepare datasets
  – Development of trial with view to sharing
  – Preparation and documentation of datasets at “end”

• Effort to support released datasets
  – None if not requested
  – Much if many secondary requests that provoke engagement
Using data in valid way & utility of new knowledge

- Trials analyzed by highly qualified teams with
  - Clear stats planning in design
  - Statistical analysis plans detailing analyses
  - Primary, secondary and exploratory, & accounting for multiple testing

- Trials teams routinely criticized if (perceived to be) doing lots of testing and subgroup analyses
  - Some ICMJE journals will not put subgroups analyses into abstracts
  - Reproducibility crisis in psychological literature

- How to protect trial integrity and integrity of secondary analysis?
Using data in valid way & utility of new knowledge

- Applicants set out analyses in advance
  - Needs to know whether data would be available?
- Applicant qualifications logged?
- “Approved researcher” status – register with relevant professional body
- Agree to repeat analyses and do, and work together if differences
- Specific training courses for new researchers seeking data for secondary use?

Some Challenges

Using data in a valid way and usefulness of new knowledge generated from data sharing

Rationale for linking to trial data and linkage to routine data for sharing

Oversight processes necessary for access to shared data

Recognition of researchers involved in creating data
2. Which data source and why

- Many applications require trial data
  - e.g. IPD meta-analysis, linkage to samples

- Anticipate proportion of applications do not need individual trials, just good, prospectively collected data
  - e.g. estimating events rates in some subset of patients
  - Trial = convenience sample set for researchers who can’t access prospective data when needed

- What if better access to routinely-collected health data (RCHD)\(^*\)?
  - Good EHR → few applications for trial data?

- Solve issues for access to RCHD at same time or separately to trials?

\(^*\)Electronic health records
Some Challenges

Using data in a valid way and usefulness of new knowledge generated from data sharing

Rationale for linking to trial data and linkage to routine data for sharing

Oversight processes necessary for access to shared data

Recognition of researchers involved in creating data
3. Oversight and timing
3. Oversight and timing

Who do data release requests go to?

Current

TSC = Trial Steering Committee

MRC CTU at UCL
3. Oversight and timing

Who do data release requests go to?

Number of Data Release Requests

Start of trial  Recruiting  Follow-up  Longterm Follow-up  Afterwards  Long Afterwards

Future Stage 1

TRIAL SPECIFIC TSC

+IDMC for interim data  +IDMC for unpublished data

UNIT DATA ACCESS COMMITTEE

TSC = Trial Steering Committee

MRC CTU at UCL
3. Oversight and timing

Who do data release requests go to?

- **1. Start of trial**
- **2. Recruiting**
- **3. Follow-up**
- **4. Longterm Follow-up**
- **5. Afterwards**

Number of Data Release Requests

Higher

Lower

Future Stage 2

- **TRIAL SPECIFIC TSC**
  - +IDMC for interim data
  - +IDMC for unpublished data

- **UNIT DATA ACCESS COMMITTEE**

- **EXTERNAL DATA ACCESS: CSDR & DATA ARCHIVE**
  - Or similar

Explicit input of trial team fades over time

TSC = Trial Steering Committee

MRC CTU at UCL
Some Challenges

Using data in a valid way and usefulness of new knowledge generated from data sharing

Rationale for linking to trial data and linkage to routine data for sharing

Oversight processes necessary for access to shared data

Recognition of researchers involved in creating data
A question of recognition
A question of recognition

Rashida Jones
imdb.com/name/nm0429069

Toby Jones
imdb.com/name/nm0429363

Chawana Jones
imdb.com/name/nm6398989/
A question of recognition

Rashida Jones
imdb.com/name/nm0429069
A question of recognition

Toby Jones

imdb.com/name/nm0429363
A question of recognition

Chawana Jones
Authorship alone does not recognise input

| Columns = 10+ editorials and 30+ papers for one clinical trial protocol |
| Rows = authors (names removed) |
| Cells = green if named author, red if not |
Proposal for recognition

• Would “film credits” for trials improve recognition of effort?

• Would “IMDb for trial engagement” allow recognition and ameliorate some anxieties about data sharing?
  – How might this differ between commercial and non-commercial settings?

• Recognition of researchers involved in creating data
  – Differences between academia and industry?

• DOI for shared datasets
  – Each version uploaded or one per trial?

• New models need buy in from:
  – Journals
  – Promotion committees

doi: 10.1056/NEJMc1707245
Two other challenges to mention quickly…

• How / when to **properly** anonymise / pseudonymise / de-link datasets
  – Dedicated training courses for statisticians and data scientists
  – Could be a specialist (and credited!) role at (larger?) organisations?

• Feedback of results to participants?
  – Getting better at this for main trial results
  – Whose responsibility after data sharing?

Summary

• Number of challenges

• Addressing them requires buy in secondary data users as well as primary trial researchers and broader trial community

• Society needs us to find ways to ensure answering the questions that matter in the best way

Some Challenges

- Using data in a valid way and usefulness of new knowledge generated from data sharing
- Rationale for linking to trial data and linkage to routine data for sharing
- Oversight processes necessary for access to shared data
- Recognition of researchers involved in creating data
Challenges and disincentives for sharing and using data: A researcher perspective

Professor Matthew Sydes
MRC CTU at UCL
Institute of Clinical Trials and Methodology
London, UK

Sharing Clinical Trial Data: Challenges and a Way Forward: A Workshop
17-Nov-2019