THE ETHICS OF DATA SHARING

2021 CSE Annual Meeting
5 May 2021, Session 5.1

Dr Trevor Lane, COPE Council
Consultant, AsiaEdit
THE ETHICS OF DATA SHARING

Agenda

• Some data problems
• COPE Core Practice: *Data & reproducibility*
• Data stewardship & Responsible data sharing
• Challenges & opportunities
• *Interactive cases:* Case 1 & Case 2
SOME DATA PROBLEMS

- Screen, detect, verify raw/original data
- Respond to allegations, investigate with institution*
- Correct/retract, train/prevent

*CLUE Guidelines
*COPE Guidelines
*RePAIR Guidelines
SOME DATA PROBLEMS

FFP
- Fabrication
- Falsification
- Plagiarism

QRP
- Questionable Research Practices
- Unauthorized data collection / use
- Underpowered, faulty methods
- Sampling / confounder bias
- Extending, ending early, censoring
- Cherry picking
- Fishing
- Trawling, dredging
- P hacking
- HARK

PROMOTING INTEGRITY IN SCHOLARLY RESEARCH AND ITS PUBLICATION

publicationethics.org
SOME DATA PROBLEMS

FFP
- Fabrication
- Falsification
- Plagiarism

QRP
- Unauthorized data collection / use
- Underpowered, faulty methods
- Sampling / confounder bias
- Extending, ending early, censoring
- Cherry picking
- Fishing
- Trawling, dredging
- P hacking
- HARK

Questionable Publishing Practices
- Poor record keeping / reporting
- Image distortion, exaggeration
- Misinterpretation, spin
- Unauthorized publication / sharing
- Hiding, withholding data / info
- Supporting predatory journals
- Salami slicing
- Redundant / duplicate publication
- Selective publication, file drawer problem
- Not correcting errors
COPE CORE PRACTICE
Data and reproducibility

"Journals should include policies on **data availability** and encourage the use of **reporting guidelines** and **registration** of clinical trials and other study designs according to standard practice in their discipline."

**Example COPE Forum cases**

- **Case 15-17**: Researcher who provided a database wants to be named an author post publication
- **Case 17-02**: Reader suspects image manipulation; author/institution says originals are lost
- **Case 11-04**: Journal editor proposes raw data should be deposited in database owned by editor
- **Case 12-31**: Journal asks author to supply sensitive data to reader, who then gets different results

https://cope.onl/data
COPE CORE PRACTICES

Policies and practices required to reach the highest standards in publication ethics:

- Allegations of misconduct
- Authorship and contributorship
- Complaints and appeals
- Conflicts of interest/Competing interests
- Data and reproducibility
- Ethical oversight
- Intellectual property
- Journal management
- Peer review processes
- Post-publication discussions and corrections
DATA STEWARDSHIP

- **Laws & regulations:**
  - Research governance, study registration, data / metadata sharing
  - Individuals’ right to privacy; exclusions for anonymized data

- **Ethics approval; Participant consent to participate** & for researcher to
  - Collect, process, store confidential data;
  - Publish certain data/images;
  - Share irreversibly anonymized, deidentified data, including future purposes (meta-analyses)

- **Ethical / legal use** of (others’) datasets, illustrations

- **Registered reports + results reported later** ([www.cos.io/initiatives/registered-reports](http://www.cos.io/initiatives/registered-reports))

- **Register clinical trial protocols + summary results** (eg, [www.clinicaltrials.gov](http://www.clinicaltrials.gov))

- **Data management plan, (Statistical) Analysis plan, Data publishing / sharing plan;** Standard operating procedures (secure data system, data entry/changes, data lock, archive)

- **Data for journal inspection; Complete reporting:** Equator Network Reporting guidelines ([www.equator-network.org](http://www.equator-network.org))
RESPONSIBLE DATA SHARING

- **Journal / funder data sharing policies** (inc. metadata, protocol, codebook, code, software, …)

- **Repositories** *(Registry of Research Data Repositories, www.re3data.org)*

- **DOIs / persistent identifiers, Data citation** *(FORCE11 Citation Principles)*

- **FAIR** (Findable, Accessible, Interoperable, Reusable) principles for digital data *(www.go-fair.org/fair-principles/)*

- **CARE** (Collective benefit, Authority to control, Responsibility, Ethics) principles for indigenous data *(www.gida-global.org/care)*

- **Data journals, (Meta)Data articles**, Protocol papers, Online protocols; **STAR, MDAR checklists**

- **Open Science / Research / Scholarship; Open Access / Data / Innovation**;
  - appropriate governance, infrastructure & standards (not allow re-identification or discrimination);
  - free access if public funding *(eg, OECD Recommendation of the Council concerning Access to Research Data from Public Funding)*

---

Publish & share data

- Replication
- Data integrity
- Data preservation
- Scientific progress
- Public trust
- Human rights

https://academic.oup.com/bioscience/article/63/6/483/226339
RESPONSIBLE DATA SHARING
Global Alliance for Genomics and Health

• **Principles:**
  o **Respect** individuals, families, and communities
  o **Advance** research and scientific knowledge

  o **Promote** health, well-being, fair distribution of benefits
  o **Foster** trust, integrity, and reciprocity

• **Core elements:**
  o **Transparency**
    (eg, purposes, processes, procedures, governance)
  o **Accountability**
    (eg, systems, access, conflicts of interest, complaints/sanctions)
  o **Engagement** with stakeholders
  o **Data quality & security**
    (eg, storage, processing, risks, lawful requests)

  o **Privacy, data protection & confidentiality**
    (eg, relevant regulations, no re-identification)
  o **Risk-benefit analysis**
  o **Recognition & attribution**
  o **Sustainability**
  o **Education & training**
  o **Data accessibility & dissemination**

---

RESPONSIBLE DATA SHARING

ICMJE (International Committee of Medical Journal Editors) (www.icmje.org)

• From 1 July 2018: data sharing statement (eg, will deidentified individual-participant data, codebook, protocol, analysis plan be shared? How & for how long?)

• Advance online clinical trial registration; From 1 January 2019: include data sharing plan

• Editors may request all authors to declare they had full access to data

• Permission & acknowledgment to reuse or publish others’ data

• Declare similar papers; encouraged to use reporting guidelines

Declaration of Helsinki (www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/)

• Public registration of human studies

• Complete, accurate public access to results (negative, inconclusive, or positive)

• Ethical reporting
RESPONSIBLE DATA SHARING
TOP (Transparency & Openness Promotion) (www.cos.io/initiatives/top-guidelines)

- TOP Factor & TOP Guidelines:
  - 8 standards: Data citation, data / code / materials transparency/sharing, design/ analysis transparency, study preregistration, analysis plan preregistration, replication
  - 3 levels: 
    1. Whether data are available & where
    2. Must be posted to repository
    3. Must be posted to repository & analyses reproduced independently

FORCE11 (Future of research communications and e-scholarship) (www.force11.org)

- Research Data Publishing Ethics working group (in collaboration with COPE)
**CHALLENGES & OPPORTUNITIES IN DATA SHARING**

**Publishing**

- **Guidelines & definitions:**
  - Type of data policy/statement; conditions?
  - Data types & formats?
  - Timing? Extent of “underlying data”; no more “(data not shown)”?

- **Repositories** or Journal / Supplementary materials
  - Assign DOIs / persistent identifiers?
  - File access, license (according to funder)?
  - Link to context (protocol, ethics, funding, COIs)?
  - Embargoes? Versions / copies?

- **Citation**, linking to data, permissions (even if CC0; no © for raw research data but © for compilations / selections / arrangements)

- **Exceptions / exemptions or controlled access:** legal, ethical, (commercial) sensitivity, safety reasons? *Did the authors have access to raw data?*

- **Processes:**
  - Check repository, FAIR/CARE, citation, ethical / legal, permissions, anonymity / deidentification (eg, BMJ checklist; US Safe Harbor Method)?
  - Data / statistics peer review?
  - Check appropriate reuse/publication of 2° data

- **Investigations**: Data specialists? Check for misconduct/QRPs/QPPs? Corrections? Retractions?

- **Training** of editors, reviewers, journal staff, authors?
CHALLENGES & OPPORTUNITIES IN DATA SHARING

Institutions / Funders (see Sherpa Juliet)

- **Mandated** data registration, sharing, venue, access, license; OA journal?
  - Level of data & timing?
  - Also for preprints / books / conferences?
  - Checking? Sanctions?

- **Data management plan**, analysis plan, publishing/sharing plan; approvals, amendments?

- **Data ownership**, authorship/contributorship
  - Author attitudes, knowledge, data literacy, digital skills? Credit / incentives for sharing?
  - Embargoes to allow author first-use, patents & prevent scooping?
  - Derivative datasets?

- **Institutional repositories**
  - Technical aspects, costs, time? Quality control, preservation, backups? DOIs?
  - Legacy format, versions, security, access levels (closed – mediated – open), curation, duration & continued participant protection?
  - Responsibilities or data transfer when authors leave? Ethics committee approval?

- **Investigations***:
  - Data protection / access / integrity personnel?
  - Misconduct investigations? If authors leave?
  - Check whole portfolio? Data Corrections / Retractions?

*CLUE Guidelines  *COPE Guidelines  *RePAIR Guidelines
CHALLENGES & OPPORTUNITIES IN DATA SHARING

Progress

• Adrian Ziderman, Michal Tal-Socher, *Prometheus* 2020;36:116-134: Data sharing policy for 150 journals, 15 disciplines:
  Biomedical (67%) > physical, social > formal > arts & humanities (27%)

• Charles Piller, *Science* 13 Jan 2020: 2 y after 2018 FDA / NIH final rule on clinical trial + data registration:
  Only ~45% of 4768 trials posted results on time, 24% posted late, 32% still *not posted*

• Valentin Danchev et al. *JAMA Netw Open*. 2021;4:e2033972: 2 y after 2018 ICMJE required data sharing statement:
  Of clinical trials in JAMA, Lancet, NEJM, only 17 / 89 (19%) declaring posting of individual-participant data did post data

• COVID-19 Clinical Research Coalition & Data Sharing Working Group

• FORCE11 (Future of research communications and e-scholarship) (www.force11.org)
• RDA, Research Data Alliance (www.rd-alliance.org)
• STM Research Data (www.stm-researchdata.org/)
• GA4GH, Global Alliance for Genomics and Health (www.ga4gh.org)
• CDISC, Clinical Data Interchange Standards Consortium (www.cdisc.org)
• TranspariMED (www.transparimed.org)
• AllTrials (www.alltrials.net)
• OpenAIRE (www.openaire.eu)
• OpenPharma (www.openpharma.blog)
INTERACTIVE CASE 1

Based on Case 10-09

• Reader notices data problems in a published article & asks author for dataset

• Author refuses; reader complains to Editor

• Editor refuses to mediate & advises the two to talk directly

Was the Editor’s action appropriate?

1. Yes; editors are not responsible for postpublication data sharing

2. No; editors should ensure that authors keep promises of data sharing

3. No; editor should ask author to check data & publish correction / retraction if needed
INTERACTIVE CASE 2

Based on Case 06-25

• Journal receives submission of clinical trial results
• Authors want to share anonymized raw data as an appendix
• Data were collected several years ago; consent was not obtained for data publication / sharing

Should the Editor allow data sharing?

1. No; there was no consent or specific ethics committee approval
2. Yes, as long as data are anonymized & deidentified
3. Yes, but only after retrospective ethics committee approval

COPE Guidance for Editors: Research, Audit & Service Evaluations: https://publicationethics.org/node/19876
Journals’ Best Practices for Ensuring Consent for Publishing Medical Case Reports: Guidance from COPE: https://doi.org/10.24318/cope.2019.1.6
Iain Hrynsztkiewicz et al, Trials, 2020;11:9
THANK YOU
Trevor Lane, Council Member, COPE
Email: trevorlane@publicationethics.org