

# Open Science

Pamela Shaw, Kim Luijken, Ingeborg Waernbaum, Anne-Laure Boulesteix

STRATOS Meeting Dec 14, 2021

# Background

Today more than ever Open Science is recognized as an important research priority

- Benefits of open science showcased during COVID-19 pandemic  
Kadakia KT et al. Leveraging open science to accelerate research. *NEJM*. 2021: 384(17):e61.
- Paul Ginsparg (Founder of ArXiv.org) and Center for Open Science inaugural recipients of Einstein Foundation Award for Promoting Quality in Research.
  - <https://www.einsteinfoundation.de/en/award/>
- Stated mission of STRATOS: *The objective of STRATOS is to provide **accessible** and accurate guidance in the design and analysis of observational studies. [Sauerbrei et al. SIM 2014]*

# There have been 2 STRATOS Break out sessions on Open Science (Nov 9, Dec 1)

Participants:

Anne-Laure Boulesteix, James Carpenter, Harbajan Chadha-Boreham, Michael Kammer, Victor Kipnis, Edwin Kipruto, Kim Luijken, Willi Sauerbrei, Pamela Shaw, Ewout Steyerberg, Ingeborg Waernbaum, Shuo Wang

The following slides represent summary points from this discussion.

# What are the components of open science and how do we achieve those?

- Open access manuscripts
- Open code
- Reproducible results
- Open data
- Transparency

# Open Access Manuscripts

- A big step to open science is making research papers open access
- Pre-print servers allow papers to be immediately accessible
- Listing paper on pre-print server does not preclude publication in peer-reviewed journal
- Deposit in open access archive (e.g PubMed) often required when one of authors is supported on public funds
- Pay journal open access fee

# Open Code and Reproducible Results

- There are many ways code can be shared
  - Code to generate research results (simulations, data example) can be part of supplemental material
  - Code can be shared on GitHub.com
  - Examples in R markdown environment
  - Some journals (e.g. *Biometrical Journal*) require depositing code as part of paper submission, which gets reviewed to ensure reproducibility
- Some potential barriers:
  - Producing well-documented code takes time
    - Planning from start helpful: annotate code, use code subheadings/separate files
  - Can be time consuming to use someone else's code
  - Running code to reproduce results can require extensive computing resources (e.g. clusters, or computers running for long periods of time)

# Open Data

- Part of reproducibility is having open access data sets
  - Several warehouses of public datasets are available
- Some medical data can't be widely distributed
  - Could produce a similar simulated dataset so code can still be provided to illustrate method
- Papers should include a data availability statement

# Transparency

- Complete transparent reporting of all analyses done as part of the study, not just which results that were selected for the manuscript. Report the pre-specified analysis plan.
- Several registries exist which encourage pre-registration of studies (including study protocols) and results
  - ISRCTN registry (<https://www.isrctn.com>)
  - Open Science Framework (<https://osf.io>)
  - PROSPERO (<https://www.crd.york.ac.uk/prospero/>)
  - ClinicalTrials.gov



# Examples

- Ingeborg Waernbaum (TG7, Simulation Panel)
- Anne-Laure Boulesteix (TG9, Simulation Panel)

# Discussion: What can STRATOS do to encourage Open Science?

- Should this be an area of STRATOS guidance?
- How can we lead by example?
- What kinds of tools can we develop?
- Are we in need of a new STRATOS Panel focused on OPEN Science?
  - Natural partners: Data Sets Panel, Publications Panel, Simulation Panel
- What do we expect of our STRATOS manuscripts?
  - Deposit on pre-print server
  - Data availability statement
  - Availability of code