DATA SHARING & TRANSPARENCY

SHARON D. YEATTS, PHD

DATA COORDINATION UNIT

MEDICAL UNIVERSITY OF SOUTH CAROLINA

DISCLOSURES

- CONSULTANT FEES FROM GENENTECH RELATED TO ROLE ON PRISMS TRIAL STEERING COMMITTEE
- INSTITUTIONAL FEES FROM BARD RELATED TO DSMB SERVICE

NIH STATEMENT ON SHARING RESEARCH DATA

- 2002: Draft statement announced
  - Comments from scientific organizations and private individuals
  - HHS published final modifications for the “Privacy Rule”, which governs how covered entities use and disclose identifiable health information
- 2003: Final statement released

NIH STATEMENT ON SHARING RESEARCH DATA

- "We believe that data sharing is essential for expedited translation of research results into knowledge, products, and procedures to improve human health. "
- "Starting with the October 1, 2003 receipt date, investigators submitting an NIH application seeking $500,000 or more in direct costs in any single year are expected to include a plan for data sharing or state why data sharing is not possible."
- "In some cases, Program Announcements (PA) may request data sharing plans for applications that are less than $500,000 direct costs in any single year. "

WHO NEEDS TO COMPLY?

- Recent funding opportunity announcements indicate that all applications, regardless of the amount of direct costs requested for any one year, should address a Data Sharing Plan.
  - NINDS Efficacy Clinical Trials
  - NIH SIREN Neurological Clinical Trials
  - NIH StrokeNet Clinical Trials and Biomarker Studies for Stroke Treatment, Recovery, and Prevention
  - NeuroNEXT Clinical Trials
  - NINDS Exploratory Clinical Trials

NIH STATEMENT ON SHARING RESEARCH DATA

- Data should be de-identified in order to protect the rights and privacy of research participants
- Timely release and sharing
  - "NIH recognizes that the investigators who collect the data have a legitimate interest in benefiting from their investment of time and effort. "
  - "Timely release and sharing" is defined as no later than the acceptance for publication of the main findings from the final data set". (OCT2017 NIH Grants Policy Statement)

CONSIDERATIONS

- Budgeting: include costs to prepare data for sharing in the grant application
- Preparing data
  - Inclusion of Common Data Elements when designing CRFs
  - De-identification
  - NIH FAQ recommends guidelines prepared by the Inter-University Consortium for Political and Social Research at the University of Michigan (2012 edition: https://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf)

COMMON DATA ELEMENTS

- Subject areas: https://www.nlm.nih.gov/cde/subject_areas_1.html
- NINDS CDE
  - NINDS 'strongly encourages' researchers to ensure compatability with CDEs
  - 18 specific diseases
  - >10000 CDEs over 550 instruments
  - 4 categories: core, supplemental highly recommended, supplemental, exploratory
  - Disease-specific core: 'gold standard' measures
- CDE folks will review CRFs for compatability prior to study start

NINDS CDE

- Search tool yields
  - CDE ID/Name/Description/Data Type
  - Permissible response values and description of each
  - Question text and instructions
  - References
- mRS
- Hospital discharge destination

DATA ARCHIVES

- NINDS: https://www.ninds.nih.gov/Current-Research/Research-Funded-NINDS/Clinical-Research/Archived-Clinical-Research-Datasets
  - Data submitted at completion of trial
- Federal Interagency Traumatic Brain Injury Research (FITBIR) Informatics System, specific to TBI
  - Data periodically submitted while trial is ongoing

DE-IDENTIFICATION

- Scramble site IDs and subject IDs while maintaining link between subjects enrolled within the same site
- Scrub text fields for identifiers (or do not include in public use data set)
- Convert date/time fields to time elapsed between randomization and event
- Add order of enrollment?

NIH plans to enhance reproducibility

Francis S. Collins and **Lawrence A. Tabak** discuss initiatives that the US National Institutes of Health is exploring to restore the self-correcting nature of preclinical research.

Instead, a complex array of other factors seems to have contributed to the lack of reproducibility. Factors include poor training of researchers in experimental design; increased emphasis on making provocative statements rather than presenting technical details; and publications that do not report basic elements of experimental design[4]. Crucial experimental design elements that are all too frequently ignored include blinding, randomization, replication, sample-size calculation and the effect of sex differences. And some scientists reputedly use a 'secret sauce' to make their experiments work — and withhold details from publication or describe them only vaguely to retain a competitive edge[5]. What hope is there that other scientists will be able to build on such work to further biomedical progress?

Reproducibility is potentially a problem in all scientific disciplines. However, human clinical trials seem to be less at risk because they are already governed by various regulations that stipulate rigorous design and independent oversight — including randomization, blinding, power estimates, pre-registration of outcome measures in standardized, public databases such as ClinicalTrials.gov and oversight by institutional review boards and data safety monitoring boards. Furthermore, the clinical trials community has taken important steps towards adopting standard reporting elements[7].

We are pleased to see that some of the leading journals have begun to change their review practices. For example, Nature Publishing Group, the publishers of this journal, announced[8] in May 2013 the following: restrictions on the length of methods sections have been abolished to ensure the reporting of key methodological details; authors use a checklist to facilitate the verification by editors and reviewers that critical experimental design features have been incorporated into the report, and editors scrutinize the statistical treatment of the studies reported more thoroughly with the help of statisticians. Furthermore, authors are encouraged to provide more raw data to accompany their papers online.

Similar requirements have been implemented by the journals of the American Association for the Advancement of Science — *Science Translational Medicine* in 2013 and *Science* earlier this month[9] — on the basis of, in part, the efforts of the NIH's National Institute of Neurological Disorders and Stroke to increase the transparency of how work is conducted[10].

REPRODUCIBILITY

- Rigor and transparency, to improve reproducibility (https://grants.nih.gov/reproducibility/index.htm)
  - Robust and unbiased experimental design, methodology, analysis, interpretation and reporting of results
  - Transparent and complete description of methodologic details
    - Blinding
    - Randomization
    - Sample size calculation
    - Analysis plan