Developing a framework for a comprehensive data sharing program

Asba Tasneem, PhD, Karen Chiswell, PhD, Brian McCourt, Matt Gross, Eric Peterson, MD, MPH, Michael Pencina, PhD

Duke Clinical Research Institute, Duke University School of Medicine, Durham, NC

SOAR
Supporting Open Access for Researchers

Duke Clinical Research Institute | FROM THOUGHT LEADERSHIP TO CLINICAL PRACTICE

INTRODUCTION

There has been a surge in calls to improve transparency in clinical research.[1-3] In response, many in academia and industry have developed platforms for sharing clinical trial data.[4]

The Duke Clinical Research Institute (DCRI) has launched Supporting Open Access for Researchers (SOAR)—a data sharing initiative that includes the DCRI, academia, and industry designed to enable clinical research data sharing for the benefit of the broader research community.

Open access is good for researchers, good for innovation, and most of all good for patients.

SOAR aims to facilitate open sharing of clinical research data with responsible researchers for verification of reported results and pursuit of interesting secondary uses of data.

METHODS

There are five key elements of the SOAR data sharing program: 1) data privacy, 2) data governance, 3) data accessibility, 4) data interoperability, and 5) data discoverability.

1. Data Privacy

To protect data privacy, de-identified and/or anonymized datasets are created per the HIPAA privacy rule while preserving data integrity. Prior to making this dataset available, Duke University Institutional Review Board approval is obtained and Data Use Agreements are created in partnership with key stakeholders. Lastly, data stewards are identified *per dataset* to facilitate data sharing.

2. Data Governance

An unbiased and comprehensive review of dataset requests is an essential element of a data sharing program. In SOAR, we established an Independent Review Committee (IRC) involving subject matter experts in patient privacy, research ethics, biostatistics, and specific clinical domains. The IRC evaluates each request, and reviews results prior to publication.

3. Data Accessibility

Data are made available to the investigator via a secured analytics environment, with access controlled at the user level and exports of patient-level data are restricted unless additional approvals are granted. The analytics platform has statistical tools to aid research analysis.

4. Data Interoperability

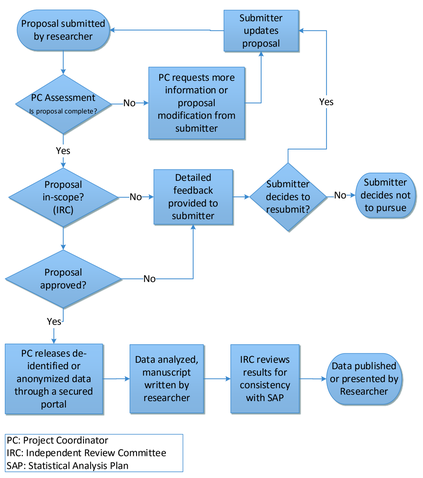
Under the SOAR data sharing program, we are looking to harmonize data using CDISC standards.

5. Data Discoverability

Data under SOAR are discoverable through a public-facing webpage on DCRI.org. After electronically signing a data use agreement, researchers can submit requests through an online data request form.

In the SOAR data sharing program we have developed a workflow to improve operational efficiency in processing data-sharing requests (Figure 1).

Figure 1. SOAR Data Sharing Workflow



RESULTS

SOAR Data Sharing Workflow

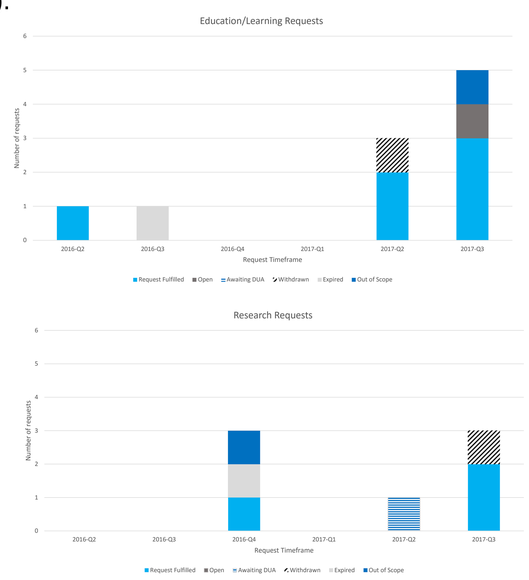
Data requests (including data use agreement and proposal) are assessed for completeness, then reviewed by the SOAR-appointed IRC. The IRC evaluates the scope of the proposal and provides recommendations to the investigator as needed. Upon approval from the IRC, the de-identified dataset(s) are provisioned for use via a secure analytics environment. The IRC also reviews analysis results prior to submission to assess concordance with the stated analysis plan (Figure 1).

Available Datasets

The DCRI is the first academic group to share its own data, starting with the Duke Databank for Cardiovascular Disease (DDCD). The DDCD is among the largest and oldest single-site cardiovascular databases in the world. The DukeCath analysis dataset, extracted from the DDCD and de-identified, includes ~150,000 cardiac catheterization procedures conducted in ~84,000 adult patients at Duke between 1985 and 2013. The DukeCath dataset is shared in two ways: 1) the de-identified research dataset (DukeCath); 2) an anonymized educational dataset (DukeCathR). These two datasets have been available to public through the SOAR program since 2016.

So far, we have received 17 proposals, out of which 10 requests were for research analysis and 7 for education and/or learning purposes (Figure 2).

Figure 2. SOAR Data Sharing Requests (2016-17)



Secured Environment for Data Analysis

The research dataset is provisioned to investigator(s) in a secured analytics environment where access is controlled at the user level and all exports of patient-level data are restricted without additional approvals. The analytics platform is equipped with statistical analysis tools to aid research analysis. We have partnered with SAS and the AHA in utilizing such analysis platforms. The Clinical Trials Data Transparency (CTDT) Multi Sponsor Environment of SAS provides SOAR researchers analytical capabilities and opportunity for collaborative analysis with industry datasets. The AHA Precision Medicine Institute platform uses the Amazon Web Services platform and provides sharing, analytical capabilities, and prospective integration with cardiovascular clinical and laboratory data contributed by other sponsors.[6]

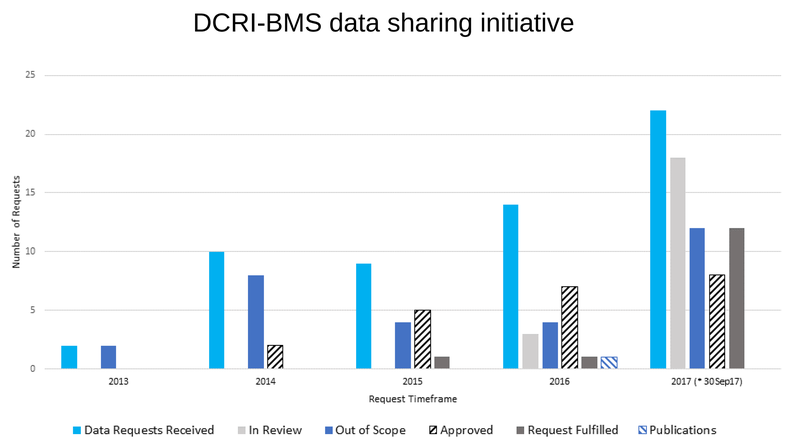
The anonymized educational dataset is available via the SAS platform or direct distribution.[6,7]

Data Partners

Under the SOAR program the DCRI has partnered with Bristol-Myers Squibb (BMS) to expand access to clinical trial information from BMS-sponsored studies. This data sharing initiative pertains to BMS-sponsored Phase I-IV interventional trials in which patients participated. This includes patient-level and study-level clinical trial data, full clinical study reports and protocols from clinical trials conducted in patients for medicines and indications approved in the United States and/or Europe. BMS uses the SOAR IRC services for reviewing in-scope proposals. Since the launch of this program in 2013 we have received 57 data requests (Figure 3).

RESULTS (cont.)

Figure 3. DCRI-BMS Data Sharing requests (2013-2017)



CONCLUSIONS

The primary purpose of the SOAR program is to facilitate open sharing of clinical research data with responsible researchers to promote open science and allow investigators to verify reported results as well as pursue interesting secondary uses of existing data. This increased transparency will inform science and improve patient care. We plan to evaluate additional clinical research datasets and share eligible datasets with researchers and educators.

The SOAR data sharing program has received requests from high school students to senior investigators, an encouraging sign for the aspirations of open science.

The SOAR program is a direct response to the widespread calls for greater transparency, data disclosure, and “open science” in clinical research and aligns with the more recent publication of the FAIR (Findable, Accessible, Interoperable, Reusable) Principles.[8]

With implementation of the new mandate to use CDISC for clinical trial submissions to FDA, there is opportunity to standardize shared data according to CDISC.[9]

REFERENCES

1. PhRMA and EFPIA: Principles for responsible clinical trial data sharing. http://phrma-docs.phrma.org/sites/default/files/pdf/PhRMAPrinciplesForResponsibleClinicalTrialDataSharing.pdf. Accessed October 30, 2017.
2. Executive Office of the President. Increasing access to the results of federally funded research. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf. Accessed October 30, 2017.
3. Ross JS, et. al. The importance of clinical trial data sharing. Circ Cardiovasc Qual Outcomes. 2012;5:238-40.
4. Navar AM, et al. Use of open access platforms for clinical trial data. JAMA. 2016;315:1283-4.
5. Pencina MJ, et al. Supporting open access to clinical trial data for researchers. Am Heart J. 2016;172:64-9.
6. DCRI and SAS open heart-disease data to researchers. https://www.dcri.org/dcri-sas-open-heart-disease-data-researchers/. Accessed October 30, 2017.
7. AHA and AWS launch cloud-based precision medicine data marketplace to accelerate scientific discovery. http://newsroom.heart.org/news/the-american-heart-association-and-amazon-web-services-launch-cloud-based-precision-medicine-data-marketplace-to-accelerate-scientific-discovery. Accessed October 30, 2017.
8. Wilkinson MD, et. al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016;3:160018.
9. CDISC. https://www.cdisc.org/. Accessed October 27, 2017.

ACKNOWLEDGEMENTS

1. We thank all members of the SOAR IRC in providing their subject matter expertize in the review process.
2. We also thank our collaborators SAS Inc., the American Heart Association, and Bristol-Myers Squibb.