The Yale Open Data Access (YODA) Project: 10 Years of Clinical Trial Data Sharing

NIH Collaboratory Grand Rounds

April 19, 2024

Joseph S. Ross, MD, MHS
Section of General Internal Medicine, School of Medicine
Center for Outcomes Research and Evaluation, Yale-New Haven Hospital

@jsross119

Potential Competing Interests

- The YODA Project is funded by research grant through Yale from Johnson & Johnson
- Research grant funding through Yale from:
  - Food and Drug Administration (FDA)
  - National Evaluation System for health Technologies (NEST)
  - NIH/NHLBI, AHRQ, PCORI
  - Arnold Ventures
- Deputy Editor at JAMA

Open Data, Open Science – Why?

Underreporting Research Is Scientific Misconduct

Iain Chalmers, FRCOG

Substantial numbers of clinical trials are never reported in print, and among those that are, many are not reported in sufficient detail to enable judgments to be made about the validity of their results. Failure to publish an adequate account of a well-designed clinical trial is a form of scientific misconduct that can lead those caring for patients to make inappropriate treatment decisions. Investigators, research ethics committees, funding bodies, and scientific editors all have responsibilities to reduce underreporting of clinical trials. An extended use of prospective registration of trials at inception, as well as benefiting clinical research in other ways, could help people to play their respective roles in reducing underreporting of clinical trials.

Source: Chalmers, JAMA 1990;263:1405-1408.

Selective Publication and Selective Reporting

- ~50% of clinical trials are never published
- Even when published:
  - Many trial publications are delayed > 2 years
  - 50% of efficacy and 65% of safety data are incompletely reported
  - Statistically significant findings more likely to be reported
  - 62% have ≥ 1 primary outcome that was changed, introduced, or omitted
- Patients and physicians frequently make treatment decisions based on only a portion of the potentially available clinical data
- Need ways to improve publication and reporting of research …

Source: Ross et al., PLOS Medicine 2009;6:e1000144; Chan et al., JAMA 2004;291:2457-2465.

the YODA PROJECT Forging a unified scientific community

Trial Registration and Results Reporting

- **1997 FDA Modernization Act, section 113, provided public access to information about ongoing clinical trials**
- **Led to creation of ClinicalTrials.gov**

ICMJE INTERNATIONAL COMMITTEE of MEDICAL JOURNAL EDITORS

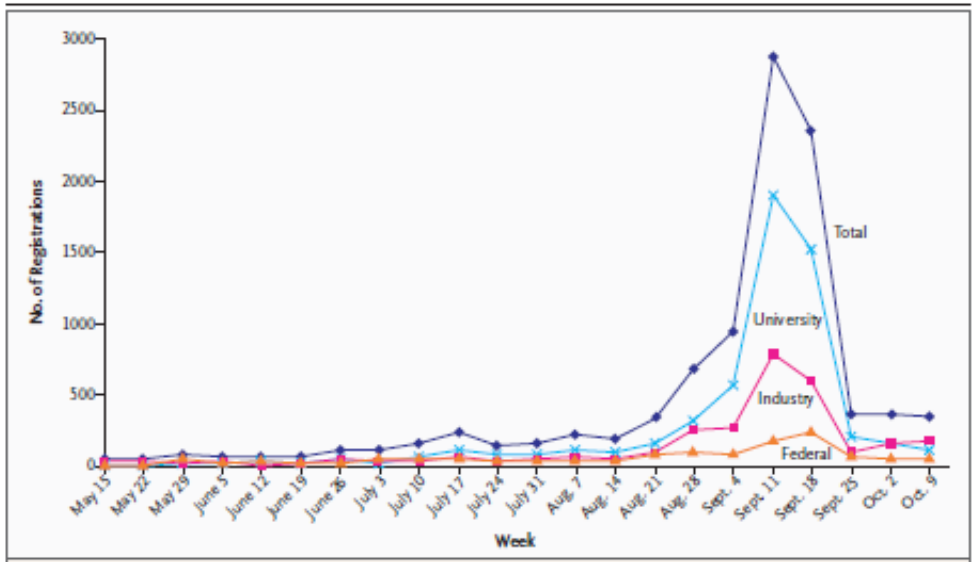


Figure 2. New Trials Registered in ClinicalTrials.gov, According to Week.

The figure shows the number of new registrations per week (beginning on the date indicated) from mid-May through early October 2005. The "Industry" category includes all commercial data providers; the "Federal" category includes the National Institutes of Health and other U.S. federal data providers; and the "University" category includes universities, foundations, and other providers.

Source: Zarin et al., NEJM 2005;353:2779-2787 | ClinicalTrials.gov.

the YODA PROJECT Forging a unified scientific community

Trial Registration and Results Reporting

- 1997 FDA Modernization Act, section 113, provided public access to information about ongoing clinical trials
- Led to creation of ClinicalTrials.gov
- **2007 FDA Amendments Act broadened scope**
  - **Expanded registry: all studies must be registered at inception**
  - **Results database: trial results uploaded within 12 months of study completion (24 if under review)**
  - **"Basic results": baseline characteristics, 1° & 2° outcomes, statistical analyses (overall & by arm)**
  - **Adverse events (serious & frequent)**

ICMJE INTERNATIONAL COMMITTEE of MEDICAL JOURNAL EDITORS

pcori PATIENT-CENTERED OUTCOME RESEARCH INSTITUTE

FDAAA

stands for

Food and Drug Administration Amendments Act

Abbreviations.com

National Institutes of Health

Turning Discovery Into Health

NHS

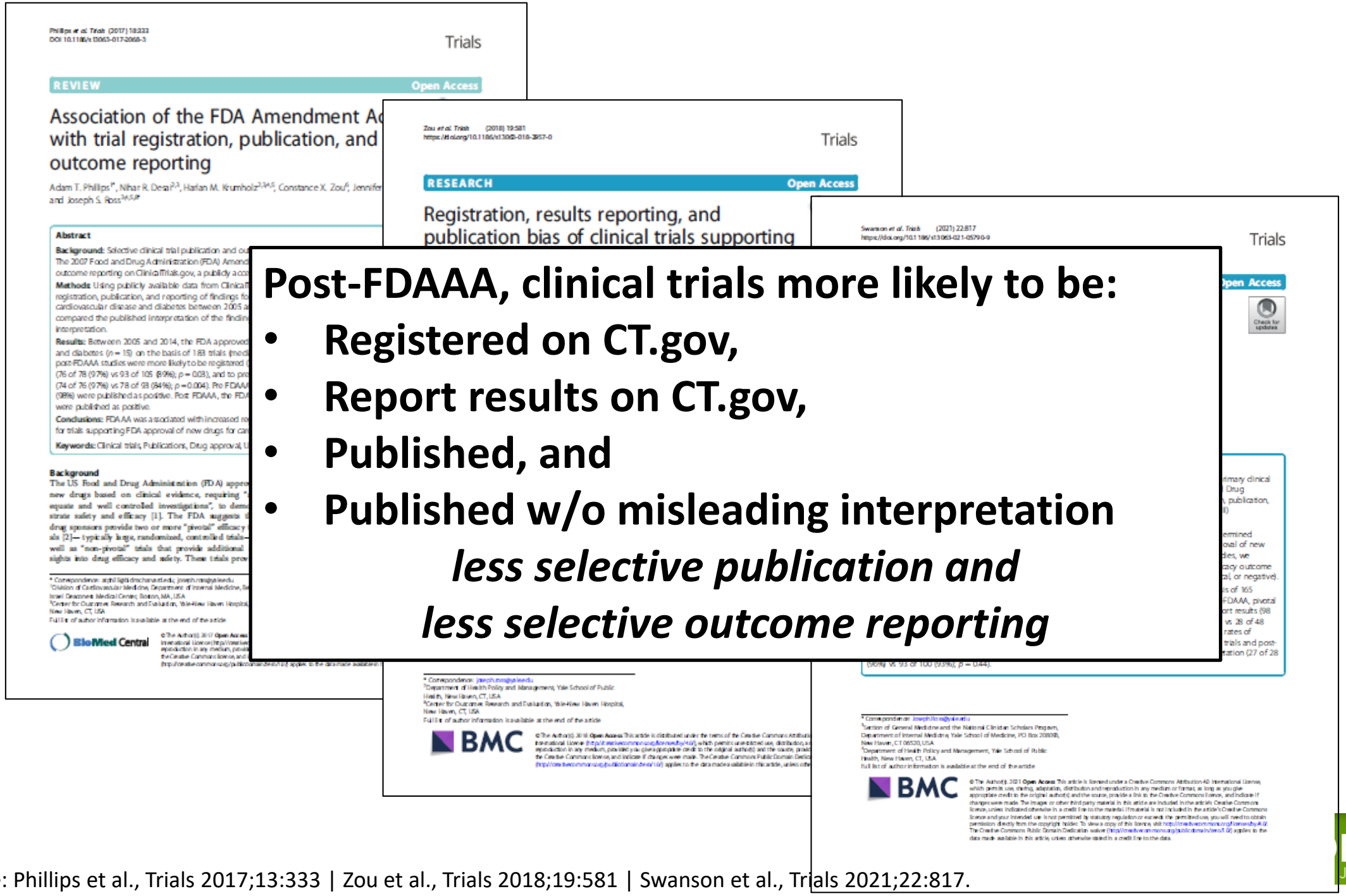
National Institute for Health Research

PhRMA

efpia

BILL & MELINDA GATES foundation

the YODA PROJECT Forging a unified scientific community



Trials

REVIEW Open Access

Association of the FDA Amendment Ac
with trial registration, publication, and
outcome reporting

Trials

RESEARCH Open Access

Registration, results reporting, and
publication bias of clinical trials supporting

Trials

Open Access

Post-FDAAA, clinical trials more likely to be:

- **Registered on CT.gov,**
- **Report results on CT.gov,**
- **Published, and**
- **Published w/o misleading interpretation**

less selective publication and
less selective outcome reporting

Source: Phillips et al., Trials 2017;13:333 | Zou et al., Trials 2018;19:581 | Swanson et al., Trials 2021;22:817.

DA JECT Forging a unified scientific community

Strengthening Science through Data Sharing

- Ensures all data can be used to inform clinical decisions
- Positions research as a public good
- Respects contributions of participants:
  - maximizing value of collected data, while
  - minimizing duplicative data collection
- Facilitates secondary studies of existing data
- Promotes transparency and reproducibility:
  - sample, design, and analysis

YODA PROJECT
Forging a unified scientific community

ICMJE INTERNATIONAL COMMITTEE of MEDICAL JOURNAL EDITORS

For clinical trials to be considered for publication:

- **Effective July 2018, manuscripts must contain a data sharing statement**
- **Trials that begin enrolling participants on or after January 2019 must include a data sharing plan in the trial's registration**

VIEWPOINT

Incentivizing a New Culture of Data Stewardship
The NIH Policy for Data Management and Sharing

For all research, funded or conducted in whole or in part by NIH, that result in the generation of scientific data:

- **Effective Jan 2023, proposals must include plans for management and sharing of all data necessary to validate and replicate research findings**

the YODA PROJECT Forging a unified scientific community

Source: Taichman et. al., Annals of Intern Med 2017;167:63-65 | Jorgenson et. al., JAMA 2021;326:2259-2260.

NIH SCIENTIFIC DATA SHARING

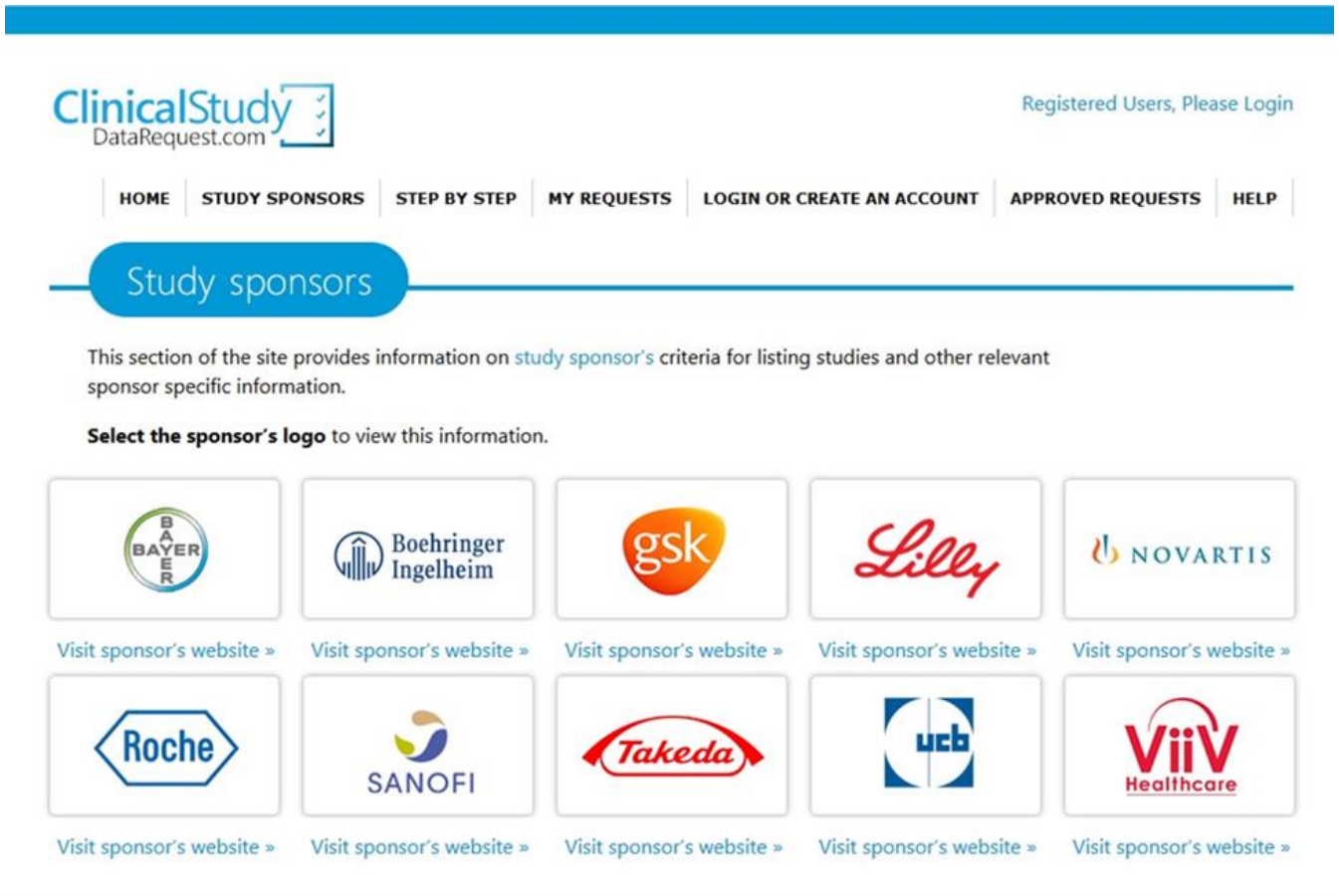
Elements to Include in a Data Management and Sharing Plan:

- Data type and amount, as well as metadata
- Related tools, software and/or code
- Standards (formats, documentation, dictionaries)
- Data preservation, access, and associated timelines
- Access, distribution, or reuse considerations
- Oversight of data management and sharing
- Budgets for allowable costs

Source: https://sharing.nih.gov/data-management-and-sharing-policy/planning-and-budgeting-for-data-management-and-sharing/writing-a-data-management-and-sharing-plan.

BioLINCC
Biologic Specimen and Data Repository Information Coordinating Center
Web Access to NHLBI Biospecimens and Data

Vivli
CENTER FOR GLOBAL CLINICAL RESEARCH DATA

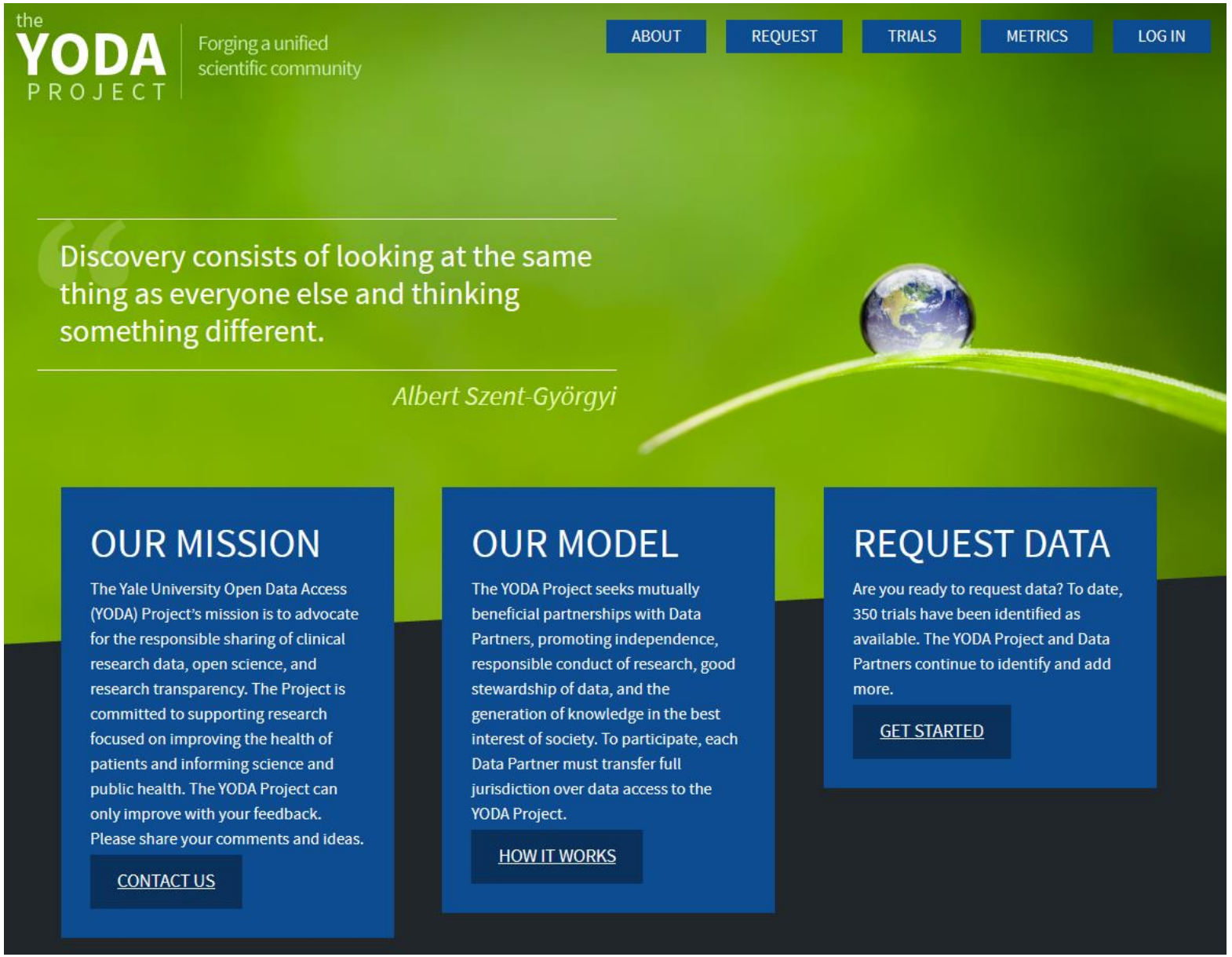


Project Data Sphere®

SOAR™

DRYAD

the YODA PROJECT
Forging a unified scientific community



the YODA PROJECT Forging a unified scientific community

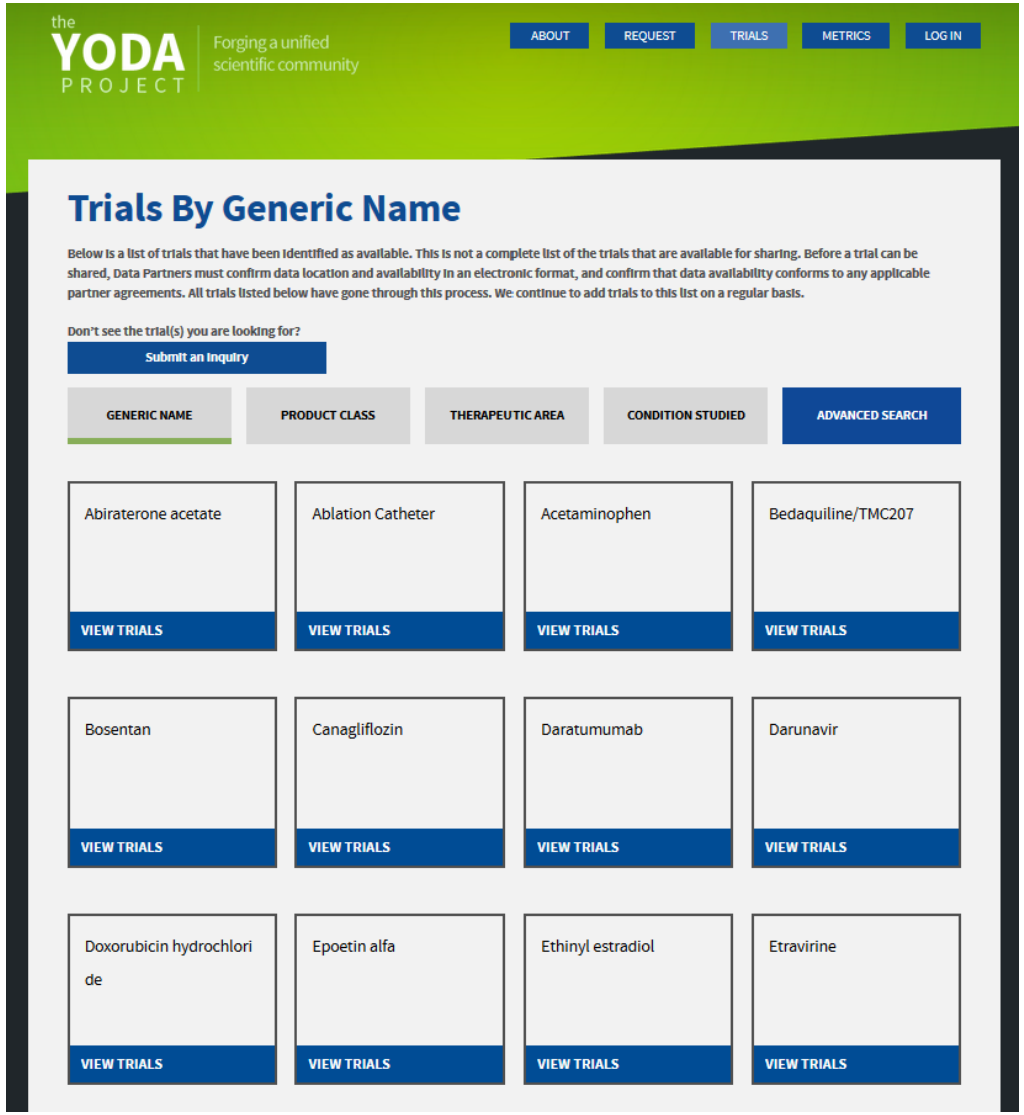
Principles of the YODA Project

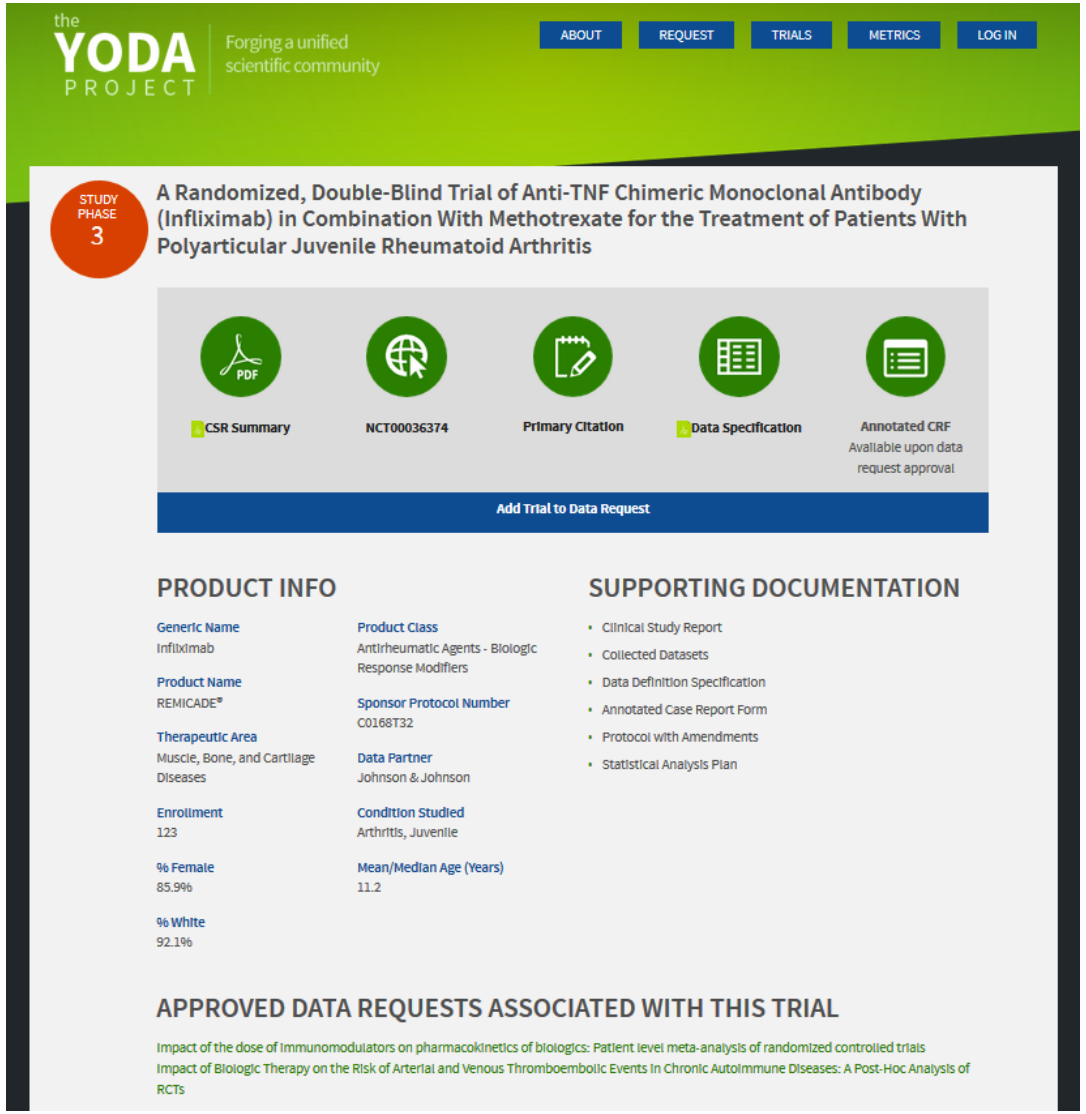
- Promote sharing of clinical research data to advance science and improve public health and healthcare
- Promote responsible conduct of research
- Ensure good stewardship of clinical research data by external investigators
- Protect rights of research participants

the YODA PROJECT Forging a unified scientific community

Johnson & Johnson Partnership

- Initiated in 2014 after proof-of-concept effort with Medtronic
- Focused on promoting and facilitating access to clinical trial data:
  - All pharmaceutical products (including legacy trials)
  - Device and diagnostic products as of 2015
  - Consumer products as of 2017
- Established data access policy and procedures, with input from Steering Committee, experts, stakeholders, and public comment





the YODA PROJECT
Forging a unified scientific community

Requests Submitted Online

- Investigator names, affiliations, funding
- Narrative summary / public abstract
- Detailed research proposal, including:
  - Project background, clear objectives
  - Trials, sample eligibility criteria, variables
  - Primary and secondary endpoints
  - Statistical analysis plan
- Project purpose (meta-analysis, validation …)
- Timeline and dissemination plan
- Data use agreement training

YODA Project Review

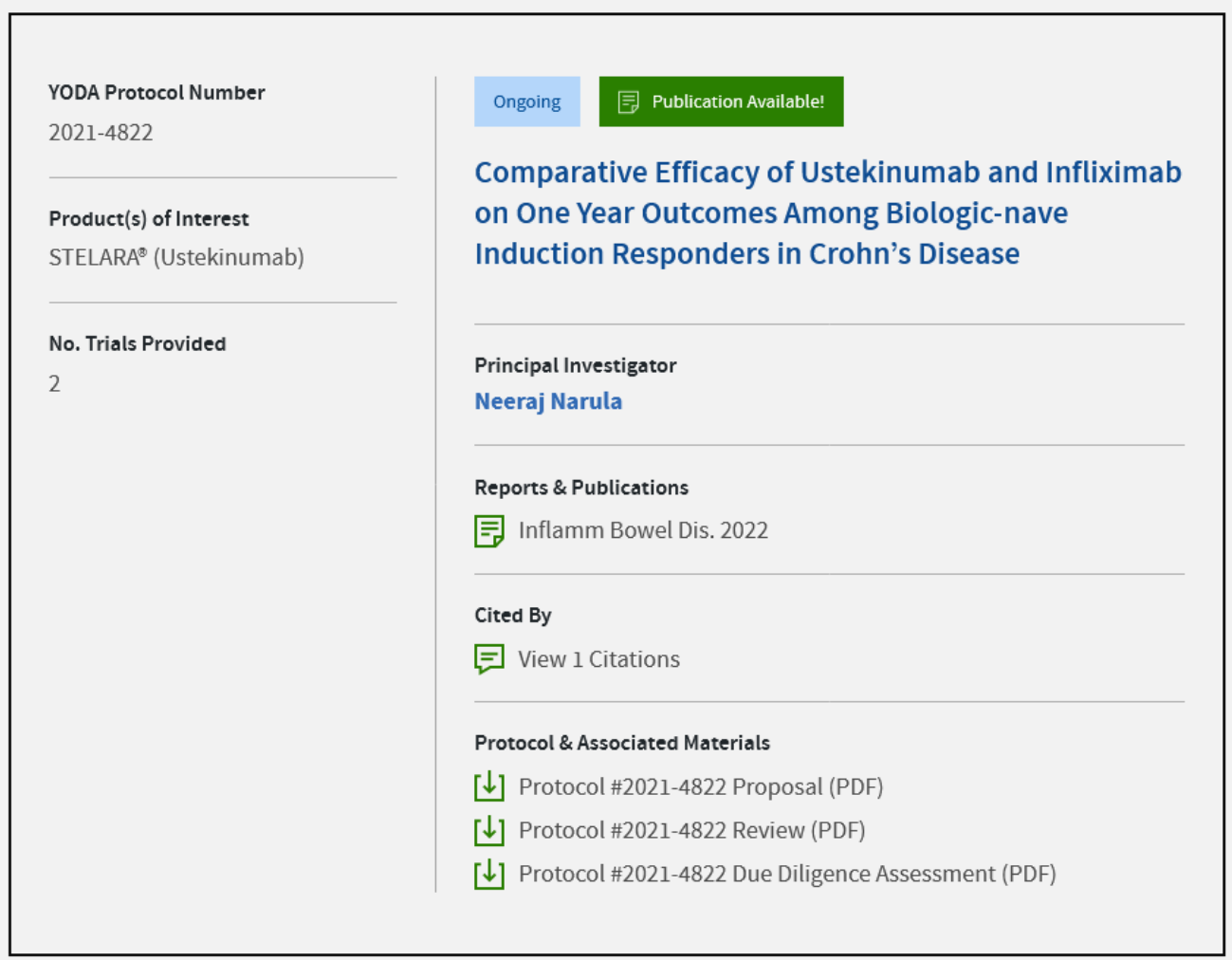
The YODA Project reviews proposals to ensure that each proposal has scientific merit, specifically verifying:

- **Scientific purpose is clearly described**
- **Data requested will be used to create or materially enhance generalizable scientific and/or medical knowledge to inform science and public health**
- **Proposed research can be pursued using the requested data**
- **Appropriateness of requested data (e.g. CSR vs IPD)**

Ensuring Data Stewardship

- Once approved, require signed DUA
- Investigators gain access to data maintained on secure platform via VPN
- Prevents re-distribution, protects patient privacy

Fostering Collaboration and Responsible Research



YODA Protocol Number
2021-4822

Product(s) of Interest
STELARA® (Ustekinumab)

No. Trials Provided
2

Ongoing | Publication Available!

Comparative Efficacy of Ustekinumab and Infliximab on One Year Outcomes Among Biologic-nave Induction Responders in Crohn's Disease

Principal Investigator
Neeraj Narula

Reports & Publications
Inflamm Bowel Dis. 2022

Cited By
View 1 Citations

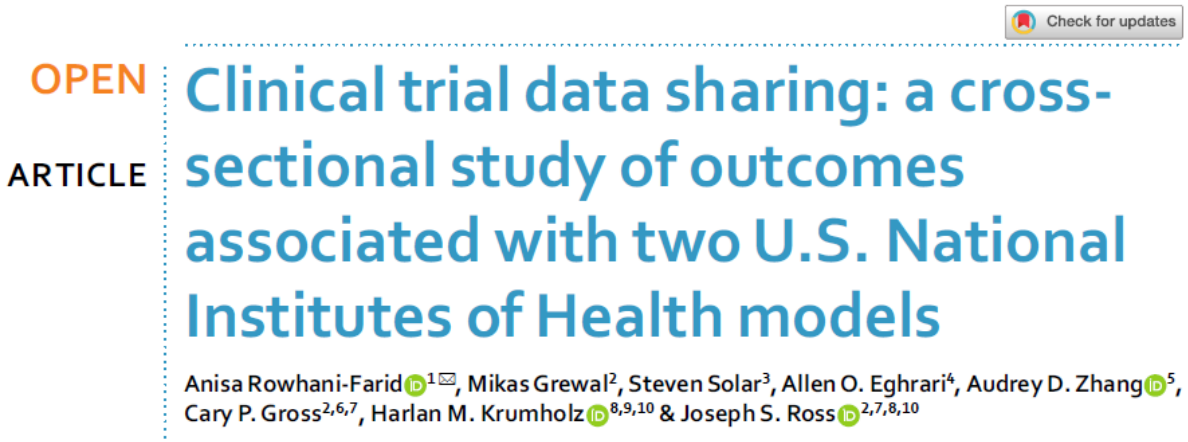
Protocol & Associated Materials
Protocol #2021-4822 Proposal (PDF)
Protocol #2021-4822 Review (PDF)
Protocol #2021-4822 Due Diligence Assessment (PDF)

the YODA PROJECT | Forging a unified scientific community

Experience so far …

- Of 459 trials currently available, 89.5% have thus far been requested
- Of 385 requests submitted, 368 (95.6%) approved, 4 (1.0%) remain under review; 11 (2.9%) withdrawn/closed, 2 (0.5%) rejected
  - Usually because data not available/cannot be adequately de-identified
- Nearly all require some administrative revision, but one-quarter required scientific revision after review for clarity
- Median number of trials per request: 3 (IQR, 1-8); 95% for IPD
- 157 manuscripts and 93 abstracts have been submitted, 119 and 89 of which have been published or presented, respectively

scientific data



- **Compared NHLBI centralized to NCI decentralized data sharing models**
- **Identified 2010-2013 trials meeting NIH data sharing criteria, matched on cost or size**
- **77 NHLBI trials, 20 (26%) shared data; 77 NCI trials, 4 (5%) shared data**
- **From the 20 NHLBI trials sharing data, we found 188 secondary internal and 53 shared data publications; for the 4 NCI trials sharing data, we found 65 secondary internal and 2 shared data publications**
- **Centralized model associated with more trials sharing data and more shared data publications**

Source: Rowhani-Farid et. al. Scientific Data 2023;10:529.

Article

CLINICAL TRIALS

Clinical Trials
1–10
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/17407745211038524
journals.sagepub.com/home/ctj
SAGE

Characteristics of available studies and dissemination of research using major clinical data sharing platforms

Enrique Vazquez[1], Henri Gouraud[2], Florian Naudet[2], Cary P Gross[3,4,5], Harlan M Krumholz[6,7,8], Joseph S Ross[3,7,8] and Joshua D Wallach[9]

	BioLINCC[a]	CSDR[c]	Project Data Sphere	SOAR-BMS[d]	Vivli[e]	YODA Project
Total number of studies listed	219	2897	154	–	(5426)[e]	395
Study design, N (%)						
Clinical trials	164 (74.9)	2876 (99.3)	154 (100.0)	–	–	394 (99.7)
Observational studies	55 (25.1)	21 (0.7)	0 (0.0)	–	–	1 (0.2)
Data sets and documents available, N (%)						
IPD (raw or analysis ready)	211 (96.3)	2884 (99.6)	154 (100)	–	–	355 (89.9)
Protocols	157 (71.7)	2773 (95.7)	84 (54.5)	–	–	359 (90.9)
Clinical study reports	4 (1.8)	2785 (96.1)	1 (0.65)	–	–	366 (92.7)
Annotative case report	171 (78.1)	2023 (69.8)	83 (53.9)	–	–	301 (76.2)
Biological specimens	49 (22.4)	0 (0.0)	0 (0.0)	–	–	0 (0.0)
Data specifications	217 (99.1)	2831 (97.7)	149 (96.8)	–	–	281 (71.1)
Data requests received, N	–	612	–	202	197	190
Data requests rejected, declined, or out of scope, N (%)	–	105 (17.5)	–	131 (64.8)	10 (5.1)	1 (0.5)
Data requests withdrawn, N (%)	–	144 (23.5)	–	9 (4.4)	24 (12.2)	18 (9.5)
Data requests in progress, N (%)	–	50 (8.2)	–	32 (15.8)	79 (40.1)	12 (6.3)
Data requests approved with contract signature, N (%)	–	313 (51.1)	–	30 (14.9)	84 (42.6)	159 (83.7)
Number of approved requests	–	313	–	30	84	159
Ongoing, data access revoked, results not reported, and unclear	–	252 (80.5)	–	24 (80.0)	78 (92.9)	107 (67.3)
In peer-reviewed journal	–	61 (19.5)	–	3 (10.0)	4 (4.8)	27 (17.0)
Other (e.g. preprint, conference abstract websites, and platform website only)	–	12 (3.8)	–	3 (10.0)	2 (2.4)	25 (15.7)

How do initially proposed aims compare with published analyses?

the YODA PROJECT Forging a unified scientific community

Source: Vazquez et. al. Clinical Trials 2021;18:657-666.

Table 3: Concordance between data publications and associated requests across all platforms.

Platform (No. pairs)	**Vivli** (n=139)	**CSDR** (n=106)	**YODA** (n=73)	**SOAR–BMS** (n=4)
Characteristics	**Publication-Request Pairs, No. (%)**	**Publication-Request Pairs, No. (%)**	**Publication-Request Pairs, No. (%)**	**Publication-Request Pairs, No. (%)**
Study objective(s)				
Fully concordant	62 (44.6)	40 (37.7)	40 (54.8)	2 (50.0)
Partially concordant	46 (33.1)	36 (34.0)	26 (35.6)	2 (50.0)
Discordant	23 (16.5)	30 (28.3)	7 (9.6)	0 (0.0)
Unclear	8 (5.8)	0 (0.0)	0 (0.0)	0 (0.0)
Trials requested and analyzed				
Fully concordant	76 (54.7)	59 (55.7)	41 (56.1)	3 (75.0)
Discordant	60 (43.2)	24 (22.6)	31 (42.5)	0 (0.0)
Greater number of trials listed in the data request	60 (43.2)	24 (22.6)	31 (42.5)	0 (0.0)
Greater number of trials listed in the publication	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Unclear number of trials in the publication	3 (2.1)	23 (21.7)	1 (1.4)	1 (25.0)
Primary endpoint(s)[a]				
Undefined primary endpoints	81 (58.3)	-	21 (28.8)	-
Methods publications	31 (38.3)	-	21 (100.0)	-
Unclear or no primary endpoints disclosed	50 (61.7)	-	0 (0.0)	-
Defined primary endpoints	58 (41.7)	-	52 (71.2)	-
Fully concordant	23(39.7)	-	23 (44.2)	-
Partially concordant	10 (17.2)	-	9 (17.3)	-
Discordant[b]	25 (43.1)	-	20 (38.5)	-
Statistical Methods				
Fully Concordant	63 (45.3)	58 (54.7)	29 (39.7)	3 (75.0)
Partially Concordant	55 (39.6)	28 (26.4)	29 (39.7)	0 (0.0)
Discordant	20 (14.4)	18 (17.0)	12 (16.4)	0 (0.0)
Unclear	1 (0.7)	2 (1.9)	3 (4.2)	1 (25.0)

Source: Vazquez et. al. under review.

2 (1.4) **19 (17.9)** **3 (4.2)** **2 (50)**

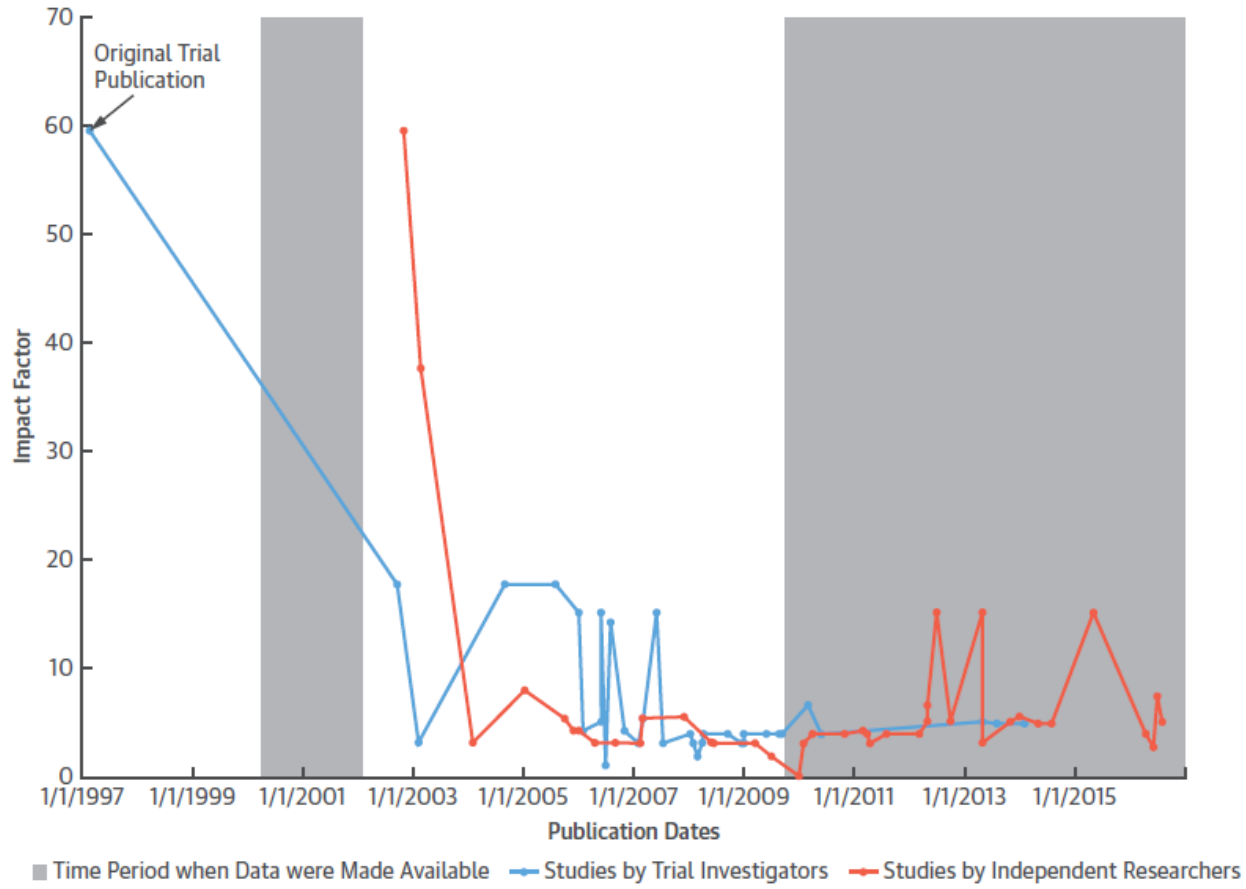
Fully concordant for all reported characteristics

Merits of Data Sharing

The Digitalis Investigation Group Trial

- **Federally funded, run from 1991 through 1995**
- **Trial published in 1997**
- **After which, no publications until 2002**
- **From 2002 through 2016, 75 studies published, 41 (55%) by independent / outside investigators**
  - **34 inside studies: 7 in high-impact, median citations of 6.8**
  - **41 outside studies: 5 in high-impact, median citations of 4.8**
- **230 variables collected, 25% reported in 1st publication, 65% as of 2016**

FIGURE 1 Timeline of the Studies Published Using the DIG Trial Data According to the Impact Factor of the Publishing Journals



DIG = Digitalis Investigation Group. April 2000 to February 2002, data available through Limited Access Datasets program. October 2009, data released on National Institutes of Health's Biologic Specimen and Data Repository Information Coordinating Center.

Source: Angraal et. al. JACC 2017;70:1825-1826.

- Similar effort under way to evaluate the YODA Project platform
- For ~400 trials used at least once:
  - When was main trial first published?
  - How many total studies published?
  - How many by trial teams and how many by independent / outside investigators?
  - Among inside and outside studies:
    - Proportion published in high-impact journals?
    - Median citation number?
    - Proportion cited in clinical practice guidelines?
    - What else?

Source: Hakimian et. al. in process.

Strengthening Science through Data Sharing

- Numerous studies that might not otherwise have been feasible to pursue, some of which have impacted health policies and guidelines
- Facilitated direct collaborations with original investigators
- Developed efficiencies (J&J now conducts all trials intending to share)
- Replication studies have supported – not undermined – original study
- No instances of patient privacy breaches
- No publications of spurious safety findings that received unwarranted attention or disrupted patient care
- No data have been used for commercial or litigious purposes

the YODA PROJECT | Forging a unified scientific community

Challenges Remain

- Broadening awareness of data availability
- Fostering expertise in using data from clinical trials (*it's complicated*)
- Making older trial data available in contemporary formats
- Adopt data standards, across sponsors, to enable meta-analyses
- Sustainable model that covers the cost of data sharing, including centralized platform (especially for NIH)
- Data Use Agreements …
- Establish standards: when should data be available, for how long, how to reward those who share data?
- Many large pharma sharing, now NIH, what about other sponsors?

www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN

Overview and experience of the YODA Project with clinical trial data sharing after 5 years

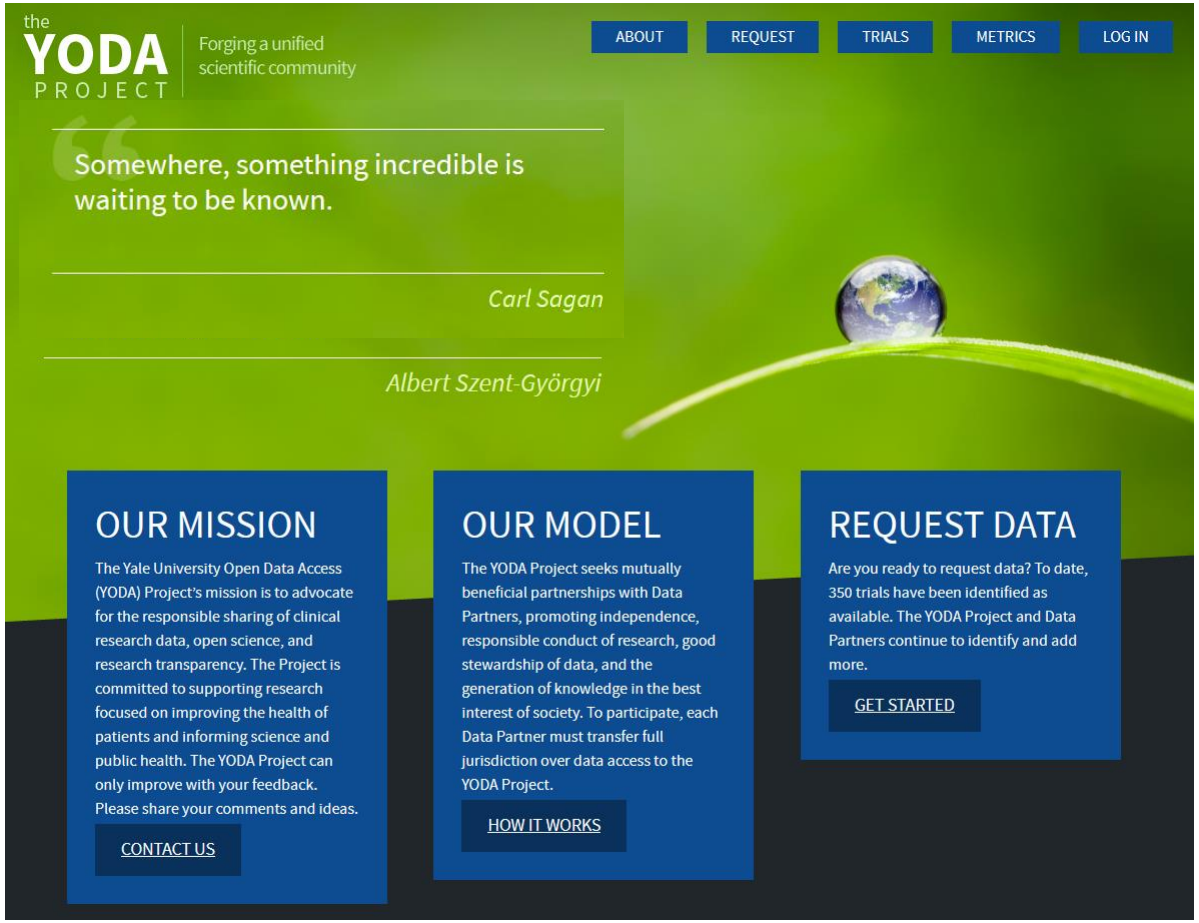
Joseph S. Ross[1,2,3,4], Joanne Waldstreicher[5], Stephen Bamford[6], Jesse A. Berlin[5], Karla Childers[5], Nihar R. Desai[4,7], Ginger Gamble[4], Cary P. Gross[1,2,4,8], Richard Kuntz[9], Richard Lehman[10], Peter Lins[5], Sandra A. Morris[5], Jessica D. Ritchie[4] & Harlan M. Krumholz[2,3,4,7]

Received: 28 March 2018
Accepted: 24 October 2018
Published: 27 November 2018

The Yale University Open Data Access (YODA) Project has facilitated access to clinical trial data since 2013. The purpose of this article is to provide an overview of the Project, describe key decisions that were made when establishing data sharing policies, and suggest how our experience and the experiences of our first two data generator partners, Medtronic, Inc. and Johnson & Johnson, can be used to enhance other ongoing or future initiatives.

the YODA PROJECT Forging a unified scientific community

Source: Ross et. al. Scientific Data 2019;5:180268.



http://yoda.yale.edu

@YODAProject **@jsross119**

joseph.ross@yale.edu

the YODA PROJECT Forging a unified scientific community