



HAL
open science

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

Enrique Vazquez, Henri Gouraud, Florian Naudet, Cary P. Gross, Harlan M. Krumholz, Joseph S. Ross, Joshua D. Wallach

► **To cite this version:**

Enrique Vazquez, Henri Gouraud, Florian Naudet, Cary P. Gross, Harlan M. Krumholz, et al.. Characteristics of available studies and dissemination of research using major clinical data sharing platforms. *Clinical Trials*, 2021, 18 (6), pp.657-666. 10.1177/17407745211038524 . hal-03335901

HAL Id: hal-03335901

<https://hal.science/hal-03335901>

Submitted on 11 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

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

Enrique Vazquez¹, Henri Gouraud², Florian Naudet², Cary P Gross^{3,4,5}, Harlan M Krumholz^{6,7,8}, Joseph S Ross^{3,7,8}, Joshua D Wallach⁹

¹Yale College, New Haven, Connecticut, USA

²Universite de Rennes, Centre Hospitalier Universitaire Rennes, Inserm, Centre d'Investigation Clinique de Rennes, Rennes, France

³Section of General Internal Medicine, Yale School of Medicine, New Haven, Connecticut, USA

⁴Cancer Outcomes, Public Policy, and Effectiveness Research (COPPER) Center, Yale University, New Haven, Connecticut, USA

⁵Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, Connecticut, USA

⁶Section of Cardiovascular Medicine, Yale School of Medicine, New Haven, Connecticut

⁷Yale-New Haven Hospital Center for Outcomes Research and Evaluation, New Haven, Connecticut, USA

⁸Department of Health Policy and Management, Yale School of Public Health, New Haven, Connecticut, USA

⁹Department of Environmental Health Sciences, Yale School of Public Health, New Haven, Connecticut, USA

Abstract

Background/Aims: Over the past decade, numerous data sharing platforms have been launched, providing access to de-identified individual patient-level data (IPD) and supporting documentation. We evaluated the characteristics of prominent clinical data sharing platforms, including types of studies listed as available for request, data requests received, and rates of dissemination of research findings from data requests.

Corresponding author: Joshua D Wallach, Yale School of Public Health, 60 College Street, Floor 4, Room 411, New Haven, CT 06510, USA, joshua.wallach@yale.edu phone: (510) 725-1445.

Author contributions: Mr. Vazquez and Dr. Wallach had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Vazquez, Naudet, Krumholz, Ross, Wallach.

Acquisition, analysis, or interpretation: All authors.

Drafting of the manuscript: Vazquez, Wallach

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Vazquez, Wallach

Supervision: Wallach.

Data sharing

Data will be shared on osf.io upon publication.

Methods: We reviewed publicly available information listed on the websites of six prominent clinical data sharing platforms: Biological Specimen and Data Repository Information Coordinating Center (BioLINCC), [ClinicalStudyDataRequest.com](https://www.clinicalstudydatarequest.com) (CSDR), Project Data Sphere, Supporting Open Access to Researchers-Bristol-Myers Squibb (SOAR-BMS), Vivli, and the Yale Open Data Access (YODA) Project. We recorded key platform characteristics, including listed studies and available supporting documentation, information on the number and status of data requests, and rates of dissemination of research findings from data requests (i.e., publications in a peer-reviewed journals, preprints, conference abstracts, or results reported on the platform's website).

Results: The number of clinical studies listed as available for request varied among five data sharing platforms: BioLINCC (n=219), CSDR (n=2,897), Project Data Sphere (n=154), Vivli (n=5426), and the YODA Project (n=395); SOAR did not provide a list of BMS studies available for request. IPD were nearly always reported as being available for request, as opposed to only Clinical Study Reports (BioLINCC, 211/219 (96.3%); CSDR, 2,884/2,897 (99.6%); Project Data Sphere, 154/154 (100.0%); and the YODA Project, 355/395 (89.9%)); Vivli did not provide downloadable study meta-data. Of 1,201 data requests listed on CSDR, SOAR-BMS, Vivli, and the YODA Project platforms, 586 requests (48.8%) were approved (i.e., data access granted). The majority were for secondary analyses and/or developing/validating methods (CSDR, 262/313 (83.7%); SOAR-BMS, 22/30 (73.3%); Vivli, 63/84 (75.0%); the YODA Project, 111/159 (69.8%)); 4 were for re-analyses or corroborations of previous research findings (CSDR, 3/313 (1.0%) and the YODA Project, 1/159 (0.6%)). Ninety-five (16.1%) approved data requests had results disseminated via peer-reviewed publications (CSDR, 61/313 (19.5%); SOAR-BMS, 3/30 (10.0%); Vivli, 4/84 (4.8%); the YODA Project, 27/159 (17.0%)). Forty-two (6.8%) additional requests reported results though preprints, conference abstracts, or on the platform's website (CSDR, 12/313 (3.8%); SOAR-BMS, 3/30 (10.0%); Vivli, 2/84 (2.4%); YODA Project, 25/159 (15.7%)).

Conclusions: Across six prominent clinical data sharing platforms, information on studies and request metrics varied in availability and format. Most data requests focused on secondary analyses and approximately one-quarter of all approved requests publicly disseminated their results. To further promote the use of shared clinical data, platforms should increase transparency, consistently clarify the availability of the listed studies and supporting documentation, and ensure that research findings from data requests are disseminated.

Keywords

Clinical trials; transparency; data sharing

Introduction

Over the past decade, the clinical research enterprise has grown exponentially. In 2020, over 34,000 clinical studies were registered with [ClinicalTrials.gov](https://www.clinicaltrials.gov)—a fifty percent increase from the number of studies registered only a decade ago.¹ Clinical studies produce vast amounts of information, including raw study data (i.e. individual patient level data [IPD]), summary level data (e.g. protocols, clinical study reports, and publications), and study documentation (e.g. data specifications). Although clinical research findings are often published in peer-

reviewed articles or shared via clinical trial registries, most clinical study data are rarely publicly available, which can hinder the ability of patients, physicians, researchers, and policy makers to make informed decisions.²³⁴ Access to clinical trial data increases the credibility of scientific research⁵ and can minimize the redundancies of duplicative experimentation, in turn reducing research costs³ and maximizing the contributions of human subjects who participate in studies.

In an attempt to facilitate access to the aggregate data and IPD from clinical studies, numerous medical organizations and funders have made data sharing recommendations or policies, including the World Health Organization,⁶ the National Academy of Medicine,⁷ the Bill and Melinda Gates Foundation,⁸ the European Clinical Research Infrastructure Network,⁹ and the U.S. National Institutes of Health (NIH).¹⁰ For instance, in 2015, the National Academy of Medicine recommended that researchers should share the analyzable data and supporting documentation from their research within 18 months of completing a clinical study.⁷ Journals have also recognized the importance of data sharing, and as of July 2018, manuscripts submitted to the International Committee of Medical Journal Editors (ICMJE) journals that report the results of clinical trials are required to include a data sharing statement outlining whether IPD and/or additional related documents will be shared.¹¹ Certain journals, including *PLOS Medicine* and *BMJ*, have implemented more stringent policies, requiring data sharing as a condition for publication of randomized controlled trials.^{12,13} Although debates surrounding the sharing of clinical data continue,¹⁴ and challenges regarding the adherence to data sharing requirements at journals remain,¹⁵ there has been growing support for data sharing, including among patients,¹⁶ researchers,⁴ and research institutions.³

In response to calls for improved transparency,¹² numerous clinical study data sharing platforms have emerged. While some platforms provide access to study data without any restrictions, others strive to promote responsible data practices and require data requestors to submit formal research proposals prior to receiving access to de-identified IPD. In one of the first partnerships of its kind, the Yale Open Data Access (YODA) Project contracted with Medtronic to share trial data from its rhBMP-2 product in 2011.¹⁷ In 2014, the YODA Project partnered with Johnson & Johnson, enabling researcher access to trial data of medical devices and pharmaceutical products.¹⁷ Since that time, other notable clinical study data sharing platforms have been launched, including [ClinicalStudyDataRequest.com](https://www.clinicalstudydatarequest.com) (CSDR), the Supporting Open Access to Researchers (SOAR) initiative run through Duke University, and Vivli (Table 1).

Despite the increased access to clinical study data, previous evaluations have suggested that individual data sharing platforms may be underutilized and largely unknown by the scientific community.¹⁸ For instance, only 15% of all cardiometabolic trials on CSDR have been accessed by investigators.¹⁹ During the first 5 years of data sharing on the YODA project platform, approximately one-third of the studies had not been requested.¹⁷ Furthermore, studies have evaluated the proportion of RCTs from certain sponsors registered at [ClinicalTrials.gov](https://www.clinicaltrials.gov) that were listed at CSDR.²⁰ In order to inform future data sharing efforts, we sought to evaluate the characteristics and use of a number of prominent clinical data sharing platforms. In particular, we determined: (1) the types of studies and data

listed on each platform; (2) the researcher and data request characteristics; and (3) the rates of dissemination (peer-reviewed publications, preprints, conference abstracts, or on the platform websites) of research using major clinical data sharing platforms.

Methods

Study design and sample section

The study protocol was pre-registered on the Open Science Framework (<https://osf.io/an29m/>).

Protocol changes

Project Genie was excluded from our evaluation after determining that it was not focused on sharing clinical study data. To streamline the data abstraction process, the therapeutic areas were updated after reviewing the eligible data sharing platforms. Many of the platforms used slightly different terms to describe the number of data requests that were approved, withdrawn, in progress, or rejected. In order to summarize this information, we standardized the status categories (i.e., approved; rejected, declined, or out of scope; withdrawn; or in progress).

Study sample and design

Two investigators (EV, JDW) conducted non-systematic Google searches to identify existing clinical study data sharing platforms. Search terms included key platform characteristics pertinent to the investigation (e.g. 'data sharing platforms', 'requestable clinical trial database', 'clinical trial data sharing platform'). Four investigators (FN, JSR, HMK, and JDW) familiar with data sharing platforms discussed potential candidates, reviewed previous evaluations,^{7,18} and selected six eligible platforms based on their size and clinical relevance: BioLINCC, CSDR, Project Data Sphere, SOAR-BMS, Vivli, and the YODA Project (Table 1). Although the Duke Clinical Research Institute (DCRI) also shares data from The Duke Databank for Cardiovascular Disease and has a catalogue of clinical research datasets on SOAR, we focused our evaluation on the partnership between DCRI and BMS.

Data Extraction

Characteristics of data sharing platforms—For each platform, one investigator (EV) independently reviewed the public access websites and recorded the dates that the platforms were established and the funders/partners. Next, we determined the number and types of studies listed as available for request (clinical trials, observational studies, and other) and the supporting documentation accessible through the platform (IPD, clinical study reports, Biological Specimen, Protocols, Annotative Case Report Forms, and Data Specifications). Within each platform, only unique studies were recorded, and listings for external data centers were excluded from the count (e.g., the YODA Project was listed on BioLINCC). The indications of each available study were also determined and categorized into four distinct groups: cardiovascular disease and diabetes, oncology and hematology, infectious disease, and other.

We reviewed whether the data sharing platforms provided publicly-available information about the data requests submitted by investigators and the status of these request (i.e., approved; rejected, declined, or out of scope; withdrawn; or in progress).

Characteristics of data requests

Two investigators (EV, HG) independently screened all approved data requests. For each platform, we screened the PDF documents describing the data requests and manually abstracted key characteristics to determine: the institution and country of origin of the primary requestor; the date of approval; and the project status (unknown, ongoing, incomplete, complete). Data request proposals were then analyzed, and projects were classified according to the purpose of the proposed research (i.e., secondary analyses and/or developing/validating methods; systematic reviews and/or meta-analyses; re-analyses or corroborations of results; or unclear).¹⁷ All data requests were reviewed by a third investigator (JDW).

Dissemination of research findings

For each data sharing platform, one author (EV) reviewed metrics for approved requests and dissemination statistics published by the platform to determine how many approved data requests were completed and had reported results in peer-reviewed publications, preprints, conference abstracts, or on the platform websites. The date of last update was recorded for all platforms.

Analyses

We conducted descriptive analyses of the main characteristics. Counts were summarized using proportions, medians, and interquartile ranges. Data were collected using Google forms and analyzed using Excel (version 16.39). All data were accessed between August and October 23rd 2020, but the date of last website update varied by platform (CSDR, 9/30/2020; SOAR-BMS, unclear; Vivli, 8/27/2020; the YODA Project, 10/23/2020). This study was conducted using publicly available, nonclinical data and did not require institutional review board approval or informed consent. This observational study was reported according to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement.²¹

Results

Characteristics of studies

As of 7 September 2020, a total of 9,091 studies were identified as listed as available for request on BioLINCC (n=219), CSDR (n=2,897), Project Data Sphere (n=154), Vivli (n=5,426), or the YODA Project (n=395) (Table 2). The SOAR platform did not provide a public list of BMS studies that were available for request. However, SOAR describes that protocols, clinical study reports, and de-identified patient level data for therapeutics and indications approved in the US and/or Europe for BMS trials completed after January 2008 can be requested. Although Vivli provided a list of 5,426 studies, the metadata describing these studies could not be downloaded. Furthermore, investigators can submit forms requesting access to the data from trials that are not listed, which are then approved

at the discretion of the data contributors. However, it is unclear how many trials could be accessed through this process. Overall, we were unable to classify the studies and supporting documentation for the SOAR-BMS and Vivli platforms.

Of the 219 studies available for request on BioLINCC, 55 (25.1%) were observational studies (Table 2). Although most study records outlined that IPD (211, 96.3%) and protocols (157, 71.7%) were available for request, only 4 (1.8%) had clinical study reports. Nearly one-quarter of the studies had biological specimens (49, 22.4%). Seventy-six (34.7%) of studies were in the therapeutic area of cardiovascular disease and diabetes.

Among the 2,897 studies available for request on CSDR, 2,876 (99.3%) were clinical trials. For the vast majority of the studies listed on CSDR, IPD (2,884, 99.6%), protocols (2,773, 95.7%), and clinical study reports (2,785, 96.1%) were listed as available for request. One-third (971, 33.5%) were in the therapeutic area of infectious diseases.

All 154 (100.0%) studies available for request via Project Data Sphere were for the control arms of clinical trials. While IPD were available for all studies (154, 100.0%), only 84 (54.5%) and 1 (0.7%) had supporting protocols and clinical study reports, respectively. All (154, 100.0%) Project Data Sphere studies were in the therapeutic area of oncology and hematology.

Among the 395 studies available via the YODA project, 394 (99.7%) were clinical trials. Nearly all of the study records listed IPD (355, 89.9%), protocols (359, 90.9%), and clinical study reports (366, 92.7%) under their supporting documentation. There was no majority therapeutic area for the studies listed on the YODA project platform.

Data request characteristics

There were 1,201 data requests listed on the CSDR (n=612), SOAR-BMS (n=202), Vivli (n=197), and YODA Project (n=190) platforms. Data request and approval metrics were not provided on the Project Data Sphere platform. Furthermore, the only metrics available for BioLINCC were reported in a published manuscript (June 2017), which focused exclusively on the sharing of biological samples, and so were not included.²²

Since the launch of CSDR in 2013, the platform has provided publicly available information on 616 data requests, of which 4 (0.65%) were duplicate (e.g., had the same title and description, but different identification numbers and dates). Among the 612 deduplicated data requests, 105 (17.5%) were classified as rejected, declined, or out of scope, 144 (23.5%) withdrawn, 50 (8.2%) in progress, and 313 (51.1%) approved (Table 3). Among the 202 requests submitted to the SOAR-BMS program as of 2013, 131 (64.8%) were classified as rejected, declined, or out of scope, 9 (4.4%) withdrawn, 32 (15.8%) in progress, and 30 (14.9%) approved. Among the 197 requests listed by Vivli since their launch in 2018, 10 (5.1%) were classified as rejected, declined, or out of scope, 24 (12.2%) withdrawn, 79 (40.1%) in progress, and 84 (42.6%) accepted. Among the 84 accepted data requests, 8 and 4 were also listed on CSDR and YODA Project platforms, respectively. Of the 190 data requests listed by the YODA Project since their launch in 2011, 1 (0.5%) was classified as rejected, declined, or out of scope, 18 (9.5%) withdrawn, 12 (6.3%) in progress, and 159

(83.7%) approved. All data on approval metrics were accessed between October 6th and 16th, 2020.

Nearly all of the 586 data requests approved by CSDR (299/313 (95.5%)), SOAR-BMS (30/30 (100%)), Vivli (78/84 (92.9%)), and the YODA Project (154/159 (96.9%)) originated from investigators from academic institutions and hospitals, who were primarily located in the USA or Canada (Table 4). Across all platforms, most data requests were for secondary analyses or developing/validating methods (CSDR, 262/313 (83.7%); SOAR-BMS, 22/30 (73.3%); Vivli, 63/84 (75.0%); YODA Project, 111/159 (69.8%)). There were only four requests listed across all platforms for studies focused exclusively on re-analyzing or corroborating previous study results (CSDR, 3/313 (1.0%); YODA Project, 1/159 (0.6%)). Although platforms did not provide information about the date of approved data requests, all data were accessed between October 6th and 16th, 2020.

Dissemination rates

Among the 313 approved data requests listed on CSDR, 61 (19.5%) had at least one corresponding publication in a peer-reviewed journal. In addition, 12 (3.8%) others had reported results through conference abstracts or on the CSDR platform (although one of these reports had broken links and one was in German); thus, 73 (23.0%) of approved data requests have reported results publicly (Table 5). Of the 30 approved data requests on SOAR-BMS, 3 (10.0%) were published in a peer-reviewed journal. In addition, 3 (10.0%) others had reported results on the SOAR-BMS platform; thus, 6 (20.0%) of approved data requests have reported results publicly. Of the 84 approved data requests on Vivli, 4 (4.8%) had a corresponding publication in a peer-reviewed journal. In addition, 2 (2.4%) others reported results on the platform website or in a conference abstract; thus, 6 (7.1%) of approved data requests have reported results publicly. Of the 159 unique approved data requests on the YODA Project, 27 (17.0%) were published in a peer-reviewed journal. In addition, 25 (15.7%) others made their results available through reports on the YODA project's platform; thus, 52 (32.7%) of approved data requests have reported results publicly.

Discussion

Our evaluation suggests that thousands of clinical studies are listed as available upon request to the scientific community on prominent data sharing platforms. However, we found differences across platforms in the amount of information describing studies and the available supporting documentation. Among the platforms providing information about the data requests that they received, the vast majority were for projects focused on utilizing shared data to conduct secondary analyses and/or validate/develop methods. Among approved data requests, nearly one-fourth had disseminated their results (peer-reviewed publications, preprints, conference abstracts, or on the platform websites) by the time of our assessment in October of 2020. As support for data sharing continues to grow, opportunities exist to increase the transparency of data sharing platforms. Furthermore, there is a need to standardize practices, increase awareness about the information available on platforms, and encourage the dissemination of findings from completed data requests.

We identified more than 9 thousand studies publicly listed as available for request on BioLINCC, CSDR, Project Data Sphere, Vivli, and the YODA Project. However, it is unclear how many of the same studies are listed on multiple data sharing platforms or how many studies can be requested but are not explicitly listed. For instance, although all studies from the Johnson & Johnson Family of Companies are listed on the YODA project platform, they are all also listed on Vivli. The same was true for certain studies on CSDR and Vivli (e.g., certain Eisai studies are listed on both platforms). Moreover, certain platforms had multiple entries for studies with the same trial identification numbers (i.e., National Clinical Trial numbers), but with slightly different names. Uncertainty also remains surrounding the actual accessibility of the studies listed on these platforms, as some data requests appear to be rejected due to a lack of available data. While the YODA Project reactively de-identifies data for sharing in response to requests,¹⁷ which has resulted in the majority of studies listed on the platform having been used by a data requestor, certain platforms (e.g., SOAR-BMS) do not list the studies that are available upon request, and appear to determine eligibility for sharing and de-identifying on a case-by-case basis. Other platforms use a proactive de-identification process, which may ultimately lead to de-identifying and making available a large number of studies never used by researchers.¹⁷ However, the ability of researchers to view all datasets that can actually be made available may help increase the number of data requests.

The vast majority of data requests approved by data sharing platforms were for secondary analysis and the development or validation of methods. Previous evaluations have highlighted the importance of using shared data for secondary analyses. For instance, in 2009, the NHLBI released the IPD from the Digitalis Investigation Group (DIG) Trial, which evaluated the effect of digoxin on mortality and morbidity in patients with heart failure. Since that time, dozens of secondary analyses have been conducted by outside investigators,²³ including one sex-based subgroup analysis demonstrating that digoxin was actually associated with a higher risk of death and a smaller reduction in the rate of hospitalization for worsening heart failure among women.²⁴ In 2017, the *New England Journal of Medicine* shared the data set underlying the Systolic Blood Pressure Intervention Trial (SPRINT) as part of its “Aligning Incentives for Sharing Clinical Trial Data” summit to encourage a wide range of novel secondary analyses. One of the key-findings from a secondary analysis of the existing data was that the originally published SPRINT study had incorrectly calculated data for the Framingham 10-year cardiovascular risk score regression model.^{25,26} In addition to data requests for secondary analyses, we identified 4 data requests focused on attempting to exclusively corroborate the findings from previously published trials. This is consistent with a previous evaluation, which found that the YODA project, SOAR, and CSDR rarely receives data requests for replication studies.¹⁸ However, the low number of requests to conduct replication studies may not be surprising, as the incentives for researchers to repeat studies are low.²⁷ Moreover, replication attempts can be challenging to conduct due to incomplete study design descriptions or datasets.

We found that one-fourth of approved data requests had findings that were disseminated via a publication in a peer-reviewed journal, preprint, conference abstract, or on the corresponding platform website. These findings demonstrate a growth in the number of publications resulting from data shared from these platforms and are a stark contrast to

a 2016 study, conducted soon after these platforms were launched,²⁸ that found only one publication had resulted from data requests to CSDR, SOAR, and the YODA Project.¹⁸ However, across all platforms, it is difficult to determine the data requests for which publications would be expected, as certain projects may be in progress or abandoned. It should be noted that challenges in dissemination may also reflect the fact that certain platforms are relatively new. Given the time that it may take to request, analyze, and publish findings,²⁹ recently founded platforms are likely to have fewer publications in peer-reviewed journals. Nevertheless, platforms have a unique opportunity to facilitate the dissemination of research findings, ensuring that updates are provided regarding project status of and results generated through data requests.

Limitations

This study had a number of limitations. First, we relied on public information from each platform's website, which meant our team was unable to access meta-data for certain platforms (i.e. SOAR's website lacked information on available studies and BioLINCC did not display request characteristics). However, our evaluation represents the best publicly available information. Second, many of the data requests outlined multiple study objectives, which made it difficult to assign requests to only one objective. While our classification system does not capture the granularity of data requests, further classification would have decreased clarity and resulted in double-counting of certain studies with multiple objectives. Third, each platform used unique terminology to describe available documentation and the status of data requests, meaning that when placed in a standardized category, certain nuances regarding a document's information were lost (e.g., manuals of procedure were classified as protocols). Furthermore, we were unable to determine why certain data requests were rejected, declined, or out of scope. Fourth, we did not account for the fact that some studies were listed on multiple platforms. However, excluding duplicate listings from the abstracted data was not possible in many cases, because our research team did not have access to the meta-data from each platform. Fifth, additional studies, supporting documentation, and data requests may have been released and/or removed since the data collection dates. Sixth, we only identified publications for data requests that were listed on the platforms. Therefore, it is possible that additional manuscripts supporting data requests were published but not linked to the platforms. Seventh, only one author (EV) characterized the dissemination characteristics for approved requests. However, these abstractions were done using a standardized approach and all uncertainties were discussed with the senior author. Lastly, we conducted non-systematic searches for clinical data sharing platforms.

While progress has been made in clinical study data sharing over the past few years,¹⁸ opportunities for improvement exist. First, there is a clear need for additional transparency. Currently, it is unclear whether all of the studies listed on these prominent platforms are actually available to investigators submitting data requests. Therefore, data sharing platforms should clearly outline the actual availability of their studies and supporting documentation. Second, there is a need to set expectations surrounding the dissemination of research findings and how long data can be used by researchers before data use agreements end, allowing researchers to access the data and check their analyses in response to comments received after results are disseminated. This information could be clarified on the pages

describing the trials available upon request. Lastly, outside of the researcher experience, there are also questions about how platforms will operate and interact with one-another in the larger data sharing ecosystem. Ensuring a congruous environment for researchers will help to further grow data sharing and make platforms more approachable to first-time users.

Conclusions

In this evaluation of six prominent clinical data sharing platforms, we found that information on studies, supporting documentation, and request metrics varied in availability and format. Most data requests were for projects focused on conducting secondary analyses or developing/validating methods. Approximately one-quarter of all approved data requests had results reported through peer-reviewed publications, preprints, conference abstracts, or on each platform website. These findings suggest that opportunities exist to standardize the information made available to researchers, further increase awareness surrounding data sharing platforms, and improve the dissemination of research using major clinical data sharing platforms.

Acknowledgments

Funding/support and role of the sponsor

Mr. Vazquez received financial support from the 2020 Yale College Dean's Fellowship and the 2020 Grace Hopper Richter Fellowship. Dr. Wallach is supported by the National Institute on Alcohol Abuse and Alcoholism of the National Institutes of Health under award 1K01AA028258. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of interest disclosure

Drs Ross and Krumholz reported being cofounders of the YODA project and Drs Gross and Wallach reported being YODA project affiliates. HMK and JSR are cofounders of the Yale Open Data Access (YODA) Project. In the past 36 months, CPG has received research funding, though Yale, from the NCCN Foundation (Pfizer/Astra-Zeneca) and Genentech, as well as funding from Johnson and Johnson to help devise and implement new approaches to sharing clinical trial data, and funding from Flatiron Inc. for travel to and speaking at a scientific conference. HMK reports that he has contracts through Yale New Haven Hospital with the Centers for Medicare & Medicaid Services to support quality measurement programs and through Yale with UnitedHealth Group to engage in collaborative research. He was a recipient of a research grant, through Yale, from Medtronic for data sharing, from the US Food and Drug Administration to develop methods for postmarket surveillance of medical devices, from Johnson & Johnson to support data sharing, and from the Shenzhen Center for Health Information for work to advance intelligent disease prevention and health promotion; he is an advisor to the National Center for Cardiovascular Diseases in Beijing, was an expert witness the Arnold & Porter Law Firm for work related to the Sanofi clopidogrel litigation, and is an expert witness for the Martin/Baughman Law Firm for work related to the Cook Select IVC filter litigation and related to C. R. Bard Recovery IVC filter litigation and for the Siegfried and Jensen Law Firm for work related to Vioxx litigation; he chairs a cardiac scientific advisory board for UnitedHealth; was a member of the IBM Watson Health Life Sciences Board; is a member of the advisory board for Element Science, the healthcare advisory board for Facebook, and the physician advisory board for Aetna; he is the cofounder of HugoHealth, a personal health information platform, and cofounder of Refactor Health, an enterprise health care artificial intelligence-augmented data management company; he is a venture partner at F-Prime. JSR is a former Associate Editor of JAMA Internal Medicine, a current Research Editor at BMJ and received research support through Yale from Johnson and Johnson to develop methods of clinical trial data sharing, from Medtronic, Inc. and the Food and Drug Administration (FDA) to develop methods for postmarket surveillance of medical devices (U01FD004585), from the Centers of Medicare and Medicaid Services (CMS) to develop and maintain performance measures that are used for public reporting, from the FDA to establish a Center for Excellence in Regulatory Science and Innovation (CERSI) at Yale University and the Mayo Clinic (U01FD005938), from the Blue Cross Blue Shield Association to better understand medical technology evaluation, from the Agency for Healthcare Research and Quality (R01HS022882), and from the Laura and John Arnold Foundation. JDW received research support through the Collaboration for Research Integrity and Transparency from the Laura and John Arnold Foundation and through the Center for Excellence in Regulatory Science and Innovation (CERSI) at Yale University and the Mayo Clinic (U01FD005938).

References

1. Trends, Charts, and Maps. U.S. National Library of Medicine, <https://clinicaltrials.gov/ct2/resources/trends> (2019, accessed 19 March 2021).
2. Wallach JD and Krumholz HM. Not Reporting Results of a Clinical Trial Is Academic Misconduct. *Ann Intern Med* 2019; 171(4): 293–294. [PubMed: 31060050]
3. Ross JS and Krumholz HM. Ushering in a New Era of Open Science Through Data Sharing: The Wall Must Come Down. *JAMA* 2013; 309(13): 1355–1356. [PubMed: 23508736]
4. Krumholz HM, Gross CP, Blount KL, et al. Sea change in open science and data sharing: leadership by industry. *Circ Cardiovasc Qual Outcomes* 2014; 7(4): 499–504. [PubMed: 24891590]
5. Gøtzsche PC. Strengthening and Opening Up Health Research by Sharing Our Raw Data. *Circ Cardiovasc Qual Outcomes* 2012; 5(2): 236–237. [PubMed: 22438464]
6. The World Health Organization. Developing Global Norms for Sharing Data and Results during Public Health Emergencies. The World Health Organization http://www.who.int/medicines/ebola-treatment/data-sharing_phe/en/ (2015, accessed 19 March 2021).
7. Institute of Medicine. *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk*. Washington, DC: The National Academies Press, 2015.
8. Kiley R, Peatfield T, Hansen J, et al. Data Sharing from Clinical Trials — A Research Funder’s Perspective. *N Engl J Med* 2017; 377(20): 1990–1992. [PubMed: 29141170]
9. Ohmann C, Banzi R, Canham S, et al. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. *BMJ Open* 2017; 7(12): e018647.
10. Hudson KL and Collins FS. Sharing and reporting the results of clinical trials. *JAMA* 2015; 313(4): 355–356. [PubMed: 25408371]
11. Taichman DB, Sahni P, Pinborg A, et al. Data Sharing Statements for Clinical Trials: A Requirement of the International Committee of Medical Journal Editors. *JAMA* 2017; 317(24): 2491–2492. [PubMed: 28586895]
12. Berlin JA, Morris S, Rockhold F, et al. Bumps and bridges on the road to responsible sharing of clinical trial data. *Clin Trials* 2014; 11(1): 7–12. [PubMed: 24408901]
13. Danchev V, Min Y, Borghi J, et al. Evaluation of Data Sharing After Implementation of the International Committee of Medical Journal Editors Data Sharing Statement Requirement. *JAMA Netw Open* 2021; 4(1): e2033972.
14. Devereaux PJ, Guyatt G, Gerstein H, et al. Toward Fairness in Data Sharing. *N Engl J Med* 2016; 375(5): 405–407. [PubMed: 27518658]
15. Taichman DB, Sahni P, Pinborg A, et al. Data sharing statements for clinical trials. *BMJ* 2017; 357: j2372. [PubMed: 28584025]
16. Mello M, Lieou V and Goodman SN. Clinical Trial Participants’ Views of the Risks and Benefits of Data Sharing. *N Engl J Med* 2018; 378(23): 2202–2211. [PubMed: 29874542]
17. Ross JS, Waldstreicher J, Bamford S, et al. Overview and experience of the YODA Project with clinical trial data sharing after 5 years. *Sci Data* 2018; 5(1): 180268.
18. Navar AM, Pencina MJ, Rymer JA, et al. Use of Open Access Platforms for Clinical Trial Data. *JAMA* 2016; 315(12): 1283–1284. [PubMed: 27002452]
19. Vaduganathan M, Nagarur A, Qamar A, et al. Availability and Use of Shared Data From Cardiometabolic Clinical Trials. *Circulation* 2018; 137(9): 938–947. [PubMed: 29133600]
20. Boutron I, Dechartres A, Baron G, et al. Sharing of Data From Industry-Funded Registered Clinical Trials. *JAMA* 2016; 315(24): 2729–2730. [PubMed: 27367768]
21. von Elm E, Altman DG, Egger M, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007; 335(7624): 806–808. [PubMed: 17947786]
22. Giffen CA, Wagner EL, Adams JT, et al. Providing researchers with online access to NHLBI biospecimen collections: The results of the first six years of the NHLBI BioLINCC program. *PLoS One* 2017; 12(6): e0178141.
23. Angraal S, Ross JS, Dhruva SS, et al. Merits of Data Sharing: The Digitalis Investigation Group Trial. *J Am Coll Cardiol* 2017; 70(14): 1825–1827. [PubMed: 28958337]

24. Rathore SS, Wang Y and Krumholz HM. Sex-Based Differences in the Effect of Digoxin for the Treatment of Heart Failure. *N Engl J Med* 2002; 347(18): 1403–1411. [PubMed: 12409542]
25. Burns NS and Miller PW. Learning What We Didn't Know — The SPRINT Data Analysis Challenge. *N Engl J Med* 2017; 376(23): 2205–2207. [PubMed: 28445656]
26. Wright JT Jr, Williamson JD, Whelton PK, et al. A Randomized Trial of Intensive versus Standard Blood-Pressure Control. *N Engl J Med* 2015; 373(22): 2103–2116. [PubMed: 26551272]
27. Wallach JD, Boyack KW and Ioannidis JPA. Reproducible research practices, transparency, and open access data in the biomedical literature, 2015–2017. *PLoS Biol* 2018; 16(11): e2006930.
28. Ross JS and Krumholz HM. Open Access Platforms for Sharing Clinical Trial Data. *JAMA* 2016; 316(6): 666.
29. Wallach JD, Egilman AC, Gopal AD, et al. Biomedical journal speed and efficiency: a cross-sectional pilot survey of author experiences. *Res Integ Peer Rev* 2018; 3(1): 1.

Table 1.

Prominent clinical data sharing repositories

Name	Program Description	Data Partners^a
Biological Specimen and Data Repository Information Coordinating Center (BioLINCC)	Data repository run by the NHLBI, a center at the NIH	Clinical study biospecimens and clinical study data from NHLBI and NIH.
ClinicalStudyDataRequest.com (CSDR)	Consortium of clinical study sponsors and funders	Astellas, Bayer, Chungai, Eisai, GSK, Novartis, Ono, Roche, Sanofi, Shionogi DSP/Sunovion, UCB, ViiV, Cancer Research UK
Project Data Sphere	Independent non-profit initiative of the CEO Roundtable on Cancer	Alliance for Clinical Trials in Oncology, Amgen, AstraZeneca, Bayer, Celgene Corporation, Clovis Oncology, ECOGACRIN Cancer Research Group, EMD Serono, Janssen Research & Development, Madrigal Pharmaceuticals, National Cancer Institute (National Institutes of Health), Pfizer, Sanofi, Takeda
Supporting Open Access to Researchers (SOAR) - Bristol-Myers Squibb	Academic-based data sharing program run through the Duke Clinical Research Institute (DCRI)	Bristol-Myers Squibb
Vivli – Center for Global Clinical Research Data	Independent non-profit organization that evolved from a project at the Multi-Regional Clinical Trials Center of Brigham and Women's Hospital and Harvard	Abbvie, Biogen, BioLINCC, Boehringer Ingelheim, Celgene, Critical Path Institute, Cure Duchenne, Daiichi-Sankyo, Doris Duke Charitable Foundation, Duke University, GSK, Harvard University, IMMPort, Johns Hopkins University, Johnson and Johnson, Lilly, Pfizer, Project Data Sphere, Regeneration, Roche, Takeda, Tempus, the Leona M. and Harry B. Helmsley Charitable Trust, UCB, and University of California San Francisco
The Yale Open Data Access (YODA) Project	Academic-based data sharing program run through Yale University	Johnson & Johnson, Medtronic, Inc, Queen Mary University of London, SI-BONE, Inc

GSK, GlaxoSmithKline; NHLBI, National Heart, Lung and Blood Institute; NIH, National Institutes of Health

^aAs of 9/2020

Table 2.Characteristics of studies available on six prominent clinical data sharing platforms^a

	BioLINCC ^b	CSDR ^c	Project Data Sphere	SOAR-BMS ^d	Vivli ^e	YODA
Total number of studies listed	219	2897	154	-	(5426) ^e	395
Study design, N (%)						
<i>Clinical trials</i>	164 (74.9)	2876 (99.3)	154 (100.0)	-	-	394 (99.7)
<i>Observational studies</i>	55 (25.1)	21 (0.7)	0 (0.0)	-	-	1 (0.2)
Datasets and documents available, N (%)						
<i>IPD (raw or analysis ready)</i>	211 (96.3)	2884 (99.6)	154 (100)	-	-	355 (89.9)
<i>Protocols</i>	157 (71.7)	2773 (95.7)	84 (54.5)	-	-	359 (90.9)
<i>Clinical Study Reports</i>	4 (1.8)	2785 (96.1)	1 (0.65)	-	-	366 (92.7)
<i>Annotative Case Report</i>	171 (78.1)	2023 (69.8)	83 (53.9)	-	-	301 (76.2)
<i>Biological Specimens</i>	49 (22.4)	0 (0.0)	0 (0.0)	-	-	0 (0.0)
<i>Data specifications</i>	217 (99.1)	2831 (97.7)	149 (96.8)	-	-	281 (71.1)
Studies per indication, N (%)						
<i>Cardiovascular Disease and Diabetes</i>	76 (34.7)	306 (10.6)	0	-	-	18 (4.6)
<i>Oncology and Hematology</i>	37 (16.9)	188 (6.5)	154 (100)	-	-	39 (9.9)
<i>Infectious Disease</i>	19 (8.7)	971 (33.5)	0	-	-	6 (1.5)
<i>Other</i>	87 (39.7)	1432 (49.4)	0	-	-	332 (84.1)

BioLINCC, Biological Specimen and Data Repository Information Coordinating Center; CSDR, [ClinicalStudyDataRequest.com](https://clinicalstudydatarequest.com); SOAR-BMS, Supporting Open Access to Researchers-Bristol-Myers Squibb (SOAR-BMS); YODA, Yale Open Data Access Project.

^aAll data were accessed on 9/7/2020.

^bThe studies displayed as available for request on the BioLINCC website originally included entries for external data-centers that can be accessed independent of the BioLINCC platform. The entries are not included in the total count.

^cThe total number of studies outlined on CSDR website appears to include duplicate entries. After deduplicating the meta data, we identified 2897 unique studies

^dStudy-specific information is not publicly available.

^eManual searches on the Vivli website, by study design categories, did not add up to the number advertised on the Vivli website (n=5459). Furthermore, the metadata describing the studies were not publicly available to verify the counts.

Table 3.Data request characteristics for data sharing platforms^a

	BioLINCC ^b	CSDR ^c	Project Data Sphere	SOAR-BMS ^d	Vivli ^e	YODA ^f
Data requests received, No.	-	612	-	202	197	190
Data requests rejected, declined, or out of scope, No. (%)	-	105 (17.5)	-	131 (64.8)	10 (5.1)	1 (0.5)
Data requests withdrawn, No. (%)	-	144 (23.5)	-	9 (4.4)	24 (12.2)	18 (9.5)
Data requests in progress, No. (%)	-	50 (8.2)	-	32 (15.8)	79 (40.1)	12 (6.3)
Data requests approved with contract signature, No. (%)	-	313 (51.1)	-	30 (14.9)	84 (42.6)	159 (83.7)

BioLINCC, Biological Specimen and Data Repository Information Coordinating Center; CSDR, [ClinicalStudyDataRequest.com](https://www.clinicalstudydatarequest.com); SOAR-BMS, Supporting Open Access to Researchers-Bristol-Myers Squibb (SOAR-BMS); YODA, Yale Open Data Access Project.

^aAll data were accessed 10/6/2020 – 10/16/2020.

^bNo data are reported on the BioLINCC website. Bio Specimen data requests reported in a manuscript published in 2017.

^cDate website last updated: 9/30/2020.

^dDate website last updated: Unclear (Q2 2020).

^eDate website last updated: 8/27/2020. The entries on the Vivli website add up to 197, although they state that there are 192. Of 84 approved requests, 4 were also listed on CSDR and 4 were also listed in the YODA project platform.

^fDate website last updated: 10/16/2020.

Table 4.

Characteristics of approved data requests, No. (%)

	BioLINCC ^a	CSDR ^b N = 313	Project Data Sphere ^c	SOAR-BMS ^d N = 30	Vivli ^e N = 84	YODA ^f N = 159
Countries of origin of primary requestor, No. (%)						
<i>USA and Canada</i>	-	133 (42.5)	-	15 (50.0)	32 (38.1)	94 (59.1)
<i>Europe</i>	-	120 (38.3)	-	11 (36.7)	33 (39.3)	48 (30.2)
<i>All Others</i>	-	60 (19.2)	-	4 (13.3)	19 (22.6)	17 (10.7)
Institution of origin of primary requestor, No. (%)						
<i>Academic Institutions and Hospitals</i>	-	299 (95.5)	-	30 (100)	78 (92.9)	154 (96.9)
<i>Other</i>	-	14 (4.5)	-	0	6 (7.1)	5 (3.1)
Purposes for each approved request, No. (%)						
<i>Secondary analyses and/or development/validation of methods</i>	-	262 (83.7)	-	22 (73.3)	63 (75.0)	111 (69.8)
<i>Systematic reviews and/or metaanalyses</i>	-	45 (14.4)	-	7 (23.3)	16 (19.0)	47 (29.6)
<i>Re-analysis/corroboration of results</i>	-	3 (1.0)	-	0	0	1 (0.6)
<i>Unclear</i>	-	3 (1.0)	-	1 (3.3)	5 (6.0)	0

BioLINCC, Biological Specimen and Data Repository Information Coordinating Center; CSDR, [ClinicalStudyDataRequest.com](https://clinicalstudydatarequest.com); SOAR-BMS, Supporting Open Access to Researchers-Bristol-Myers Squibb (SOAR-BMS); USA, United States of America; YODA, Yale Open Data Access Project.

^aInformation not accessible on BioLINCC website.

^bDate website accessed: 10/6/2020.

^cInformation not accessible on the Project Data Sphere website.

^dDate website accessed: 10/6/2020.

^eDate website accessed: 10/6/2020. Of 84 approved requests listed on Vivli, 4 were also listed on CSDR and 4 were also listed in the YODA project platform.

^fDate website accessed: 10/16/2020.

Table 5.Dissemination of results from completed data requests, No. (%)^a

	BioLINCC ^b	CSDR ^c	Project Data Sphere	SOAR-BMS ^d	Vivli ^e	YODA ^f
Number of approved requests	-	313	-	30	84	159
<i>Ongoing; data access revoked; results not reported; unclear</i>	-	252 (80.5)	-	24 (80.0)	78 (92.9)	107 (67.3)
<i>In peer-reviewed journal</i>	-	61 (19.5)	-	3 (10.0)	4 (4.8)	27 (17.0)
<i>Other (e.g., preprint, conference abstract websites, platform website only)</i>	-	12 (3.8)	-	3 (10.0)	2 (2.4)	25 (15.7)

BioLINCC, Biological Specimen and Data Repository Information Coordinating Center; CSDR, ClinicalStudyDataRequest.com; SOAR-BMS, Supporting Open Access to Researchers-Bristol-Myers Squibb (SOAR-BMS); YODA, the Yale Open Data Access Project.

^aAll data were accessed on 10/23/2020.

^bNo identifiable information.

^cDate website last updated: 9/30/2020. One of the available website reports had a broken link. One report was not in English.

^dDate website last updated: Unclear.

^eDate website last updated: 8/27/2020. Of 84 approved requests listed on Vivli, 4 were also listed on CSDR and 4 were also listed in the YODA project platform.

^fDate of last data update on website: 10/23/2020.