Published in final edited form as:

N Engl J Med. 2017 May 11; 376(19): 1849–1858. doi:10.1056/NEJMs1603542.

Use of the National Heart, Lung, and Blood Institute Data Repository

Sean A. Coady, M.S., M.A., George A. Mensah, M.D., Elizabeth L. Wagner, M.P.H., Miriam E. Goldfarb, B.S., A.S.N., Denise M. Hitchcock, B.S., and Carol A. Giffen, D.V.M.

Division of Cardiovascular Sciences (S.A.C.), the Center for Translation Research and Implementation Science (G.A.M.), and the Division of Blood Diseases and Resources (E.L.W.), National Heart, Lung, and Blood Institute, Bethesda, and Information Management Services, Calverton (M.E.G., D.M.H., C.A.G.) — both in Maryland

Abstract

Background—Research on data sharing from clinical trials has focused on elucidating perceptions, barriers, and attitudes among trialists and study participants with respect to sharing data. However, little information exists regarding utilization or associated publication of articles once clinical trial data have been widely shared.

Methods—We analyzed administrative records of investigator requests for data access, linked publications, and bibliometrics to describe the use of the National Heart, Lung, and Blood Institute data repository.

Results—From January 2000 through May 2016, a total of 370 investigators requested data from 1 or more clinical trials. Requests for trial data have been increasing, with 195 investigators (53%) initiating requests during the last 4.4 years of the study period. The predominant reason for requesting data was post hoc secondary analysis of new questions (72%), followed by analytic or statistical approaches to clinical trials (9%) and meta-analyses or pooled study research (7%). Of 172 requests with online project descriptions, only 2 requests were initiated for reanalysis of primary-outcome findings. Data from 88 of 100 available clinical trials were requested at least once, and the median time from repository availability to first request was 235 days. A total of 277 articles were published on the basis of data from 47 trials. Citation metrics from 224 articles indicated that half of the publications have cumulative citations that rank in the top 34% normalized for subject category and year of publication.

Conclusions—Demand for trial data for secondary analysis has been increasing. Requesting data for the a priori purpose of reanalysis or verification of original findings was rare.

The National Heart, Lung, and Blood Institute (NHLBI) of the National Institutes of Health (NIH) initiated a formal data repository in 2000 to facilitate the sharing of data from clinical trials and observational studies with the general scientific community. The goal of the data

Address reprint requests to Mr. Coady at the National Heart, Lung, and Blood Institute, 6701 Rockledge Dr., Rm. 10200, Bethesda, MD 20892, or at coadys@nhlbi.nih.gov.

The views expressed in this article are those of the authors and do not necessarily represent those of the National Heart, Lung, and Blood Institute, the National Institutes of Health, or the U.S. Department of Health and Human Services.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

repository is to maximize the research value of NHLBI-supported clinical studies by providing data to the widest possible audience of investigators. Overviews of the NHLBI data repository have been published previously.^{1,2} Briefly, large clinical trials and observational studies that are funded by the NHLBI deposit a copy of their deidentified, individual-level data sets in the repository after a proprietary-use period (approximately 2 years). Investigators requesting data from the repository provide a basic description of their project, provide evidence of approval by the institutional review board or ethics committee at the participating site or certification that their project is exempt from such approval, and agree to the terms and conditions of a data-use agreement. The data-use agreement is valid for 3 years, after which the researcher must renew the agreement or confirm that the data have been permanently destroyed. In October 2009, the Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC)³ began to coordinate the activities of the NHLBI biorepository and data repository.

Much of the published research on data sharing has focused on surveys of trialists and study participants to elucidate barriers to sharing, attitudes toward sharing, sharing practices, and participant perspectives.⁴⁻¹¹ However, there is little information regarding outcomes such as utilization and associated publication of articles once clinical trial data have been released for wide sharing. Sharing data from clinical trials has received particular attention recently owing to the publication of the Institute of Medicine report on the responsible sharing of clinical trial data,¹² the International Committee of Medical Journal Editors editorial on the rapid sharing of individual-level data from clinical trial publications,¹³ and the National Library of Medicine report on incorporating the sharing of individual participant data in clinical trial reporting.¹⁴

Preparing data for wide release is a nontrivial activity. Generally, data and documentation were prepared by the study submitting the data; however, in some instances, BioLINCC prepared the data. In instances in which the data preparation was done by BioLINCC, the effort to prepare the data and documentation for wide release ranged from 85 to 350 full-time-equivalent (FTE) hours, depending primarily on the quality of the documentation rather than the complexity of the data. Review of data and documentation already prepared by the study investigators has generally required approximately 80 hours of FTE effort. This process involves initial deidentification checks, interaction with data preparers to resolve questions or discrepancies, remediation of documents for the website to be compliant with Section 508 (accessible to persons with disabilities) standards, and composition of the website description. However, this phase can also vary widely. Furthermore, replication of the primary results as a check on the system process can add an additional 70 FTE hours to the review.

Given the level of effort needed to prepare data for release, host the data, and monitor use, an understanding of the benefits of wide sharing of trial data is needed. In this article, we describe the use of clinical trial data from the NHLBI data repository. Specifically, this report describes trends in the number of data requests, characteristics of investigators who requested data, and the reasons for requesting data. In addition, this report describes the time from data availability to first request and first publication and describes the number and citation metrics of resulting publications. Although this article is focused on clinical trial

data shared through the repository, data related to the sharing of observational study data are also presented to provide context and further perspective on widely shared data.

Methods

Coding of Administrative Records

Detailed methods for coding data-requester characteristics and linkage to publications are available in the Supplementary Appendix, available with the full text of this article at [NEJM.org](https://www.nejm.org). In brief, data-requester characteristics were derived from administrative records collected as part of program management. The institution of each data requester was classified as an academic, nonprofit, or for-profit entity. Requests were categorized into a research area on the basis of the study data requested. A listing of available clinical trials (as of May 31, 2016) and the classification scheme is provided in the Supplementary Appendix.

The primary purpose of each request was based on project titles (beginning in 2007) and categorized as follows: new question, defined as a secondary analysis designed to explore associations, prognostic factors, subgroup analyses, or similar issues; meta-analysis or pooled study, defined as a formal meta-analysis of individual participant data, combined study analysis, or consortium of studies with participant-level data; statistical methods, defined as a project focused on the development and testing of new statistical approaches; clinical trial methods, defined as a project examining statistical methods or analytic approaches that are generalizable to all or specific types of clinical trials; and other projects, examples of which include pilot data for a subsequent grant submission, simulation studies, and development of prediction equations. In addition, project plans were searched for any mention of a reanalysis of previously published findings; however, online project descriptions (as opposed to an uploaded document) have only recently been required as part of the data-request process. Therefore, this search was limited to 172 of 419 requests processed by BioLINCC. Access to external funding and funding source were reported by the data requesters at the time of the request.

Characteristics of clinical trials that are shared through the data repository (year trial started, intervention type, and purpose) were abstracted from ClinicalTrials.gov. If the information was not reported, the data were extracted from the study protocol or a study manual.

Publication Citation Metrics

Citation metrics for publications that are linked to data requested from the repository were derived from the Thomson Reuters InCites database, a research and bibliometric assessment tool providing citation-based metrics for most articles indexed by the Web of Science Core Collection. The citation metric that was used was the citation percentile, which was defined as the top percentile of cumulative citations that an article has generated as compared with all other articles published in the same subject category and year. Thus, the lower the percentile for a publication, the more citations it has received relative to other publications, normalized for the subject category and publication year.

Time Frames for Data Use and Associated Publications

Data requests, available trial data sets, and publication results were those known as of May 31, 2016. Years were grouped into 4-year blocks (with 2016 combined into the 2012+ year group) to simplify the presentation of trends. Citation metrics were limited to articles published before May 31, 2015, to allow at least 1 year for the accumulation of citations.

Results

Repository Use by External Investigators

A total of 1116 data requests were initiated by 844 investigators, and 2104 study databases were provided to investigators initiating data requests. In terms of provided databases, observational study data historically exceeded those for clinical trials (1237 of 2104 study databases provided [59%]); however, since 2014, delivery of clinical trial databases outpaced that for observational studies (350 of 635 study databases provided [55%]). The cumulative number of provided clinical trial databases (excluding repeated deliveries of the same study to the same investigator), along with the cumulative number of available clinical trials, are shown in Figure 1A, and the results for observational study data are shown in Figure 1B. The rate of acquisition of clinical trials available for request has increased over time, with 16% of the total available trials being released during the 2000–2003 period and 37% of trials released since 2012. However, 4% of all requests for trial data occurred during the 2000–2003 period, and 57% of trial data requests have occurred since 2012 (4.4 years), indicating an increasing demand for trial data that has outpaced acquisition. In contrast, demand for observational data has increased in a pattern more directly proportional to time.

Characteristics of investigators and requests that included at least one clinical trial are shown in Table 1. Approximately 25% of investigators requested both clinical trial data and observational study data; however, the percentage that simultaneously requested observational study data declined over time. Most requesters were from academic institutions; however, requesters from nonprofit and for-profit institutions represent a considerable proportion of the investigator pool.

Most requests were for trials of prevention or treatment of cardiovascular disorders; however, 22% of requesters asked to obtain data on lung disorders (excluding asthma), which represented 17% of the trial portfolio. Among data requests with a coded purpose, 72% of requests were initiated to address a new question or hypothesis, 7% to perform a meta-analysis or combined study analysis, 2% to test statistical methods, 9% to investigate methods relevant to clinical trials, and 9% for other reasons. Defining a reanalysis as a project in which the aim is to critically reassess the primary-outcome findings of the trial revealed only two requests in which the available description suggested a reanalysis. Both of the reanalyses examined prespecified interactions between the intervention and a specific cofactor with the primary or secondary outcome. Both of these reanalyses were published, and neither disputed the published primary result. Projects that focused on examination of specific subgroups or new statistical approaches were not considered to be reanalyses. Among requesters for which there was information on external funding, approximately two thirds reported no external funding at the time of the initial request. A total of 45 of 370

requesters (12%) were from outside the United States or Canada (Fig. 2, and see the interactive graphic, available at [NEJM.org](https://www.nejm.org)), and 111 (30%) published at least one article based on trial data obtained from the repository.

Technical support for investigators who request data is an important, often overlooked, consideration for widely shared data. Once data have been released for wide sharing, there is an expectation of access to a reasonable level of support. Although support for investigators who have received a data set was not specifically tracked, investigators who receive data through BioLINCC can post questions through the comments area of their data-request page. An examination of 238 “incident” (investigator's first request for data) requests for clinical trial data revealed that 54 investigators (23%) had questions related to the data, documentation, access to data not included in the download, download issues, or data-format issues.

Use Characteristics According to Clinical Trial

As of May 31, 2016, the repository contained data from 100 studies initiated between 1972 and 2010, with a cumulative total of nearly 350,000 participants. More than half of the trials dealt with cardiovascular conditions, 43% used a drug intervention, and 65% were treatment trials (Table 2). On the basis of a Poisson regression model, factors associated with higher rates of access requests included trials with a larger size (rate ratio, 1.02 for each additional 1000 participants; $P < 0.001$), more recent completion (rate ratio, 1.05 per calendar year of project completion; $P < 0.001$), a focus on prevention (rate ratio, 1.42 relative to treatment trials; $P < 0.001$), and a focus on lung disorders (rate ratio, 1.30 relative to cardiovascular trials; $P = 0.008$) (Table S1 and Fig. S1 in the Supplementary Appendix). Kaplan–Meier survival curves were used to examine the time from release of a study data set in the repository to the date of the first request for that data set (Fig. S2 in the Supplementary Appendix). A total of 88 trial data sets were requested at least once, and 62% were first requested within a year after repository release. The median time until the first request was 235 days (range, 3 days to 11.3 years). By comparison, observational studies in the repository had a median time until the first request of 204 days (range, 4 days to 5 years).

Kaplan–Meier curves for the time from study release to first publication for both clinical trials and observational studies are shown in Figure 3. The time until the publication of the first article associated with a request was similar for clinical trials and observational studies ($P = 0.96$ by log-rank test). Within 1 year after availability in the repository, 4% of clinical trials and 3% of observational studies had a first publication; at 5 years, 35% of clinical trials and 48% of observational studies had at least one publication. The median time to first publication was 7.2 years for clinical trials and 9.8 years for observational studies.

Publication Metrics

A total of 711 journal publications from 293 requesters were based on repository data. Among the requesters who published journal articles, 111 published 277 articles that included data from at least one clinical trial in the repository. The distribution of citation percentiles is shown in Figure S3 in the Supplementary Appendix. A total of 243 articles had a publication date before May 31, 2015, and 224 (92%) were linked to the InCites database.

For comparison, the citation percentile distribution of publications that used repository observational study data (357 of 393 [91%]) as well as a 10% random sample of all NHLBI-supported articles published during the same period (January 1, 2000, through May 31, 2015; 16,856 articles) are also shown. Approximately 23% of the articles that used clinical trial data from the repository had cumulative citations that ranked in the top 10% of all articles normalized for the publication year and subject category, and the median citation percentile was 34.0. Similarly, the median percentile was 28.3 for articles using observational study data from the repository and 29.0 for the NHLBI 10% random sample. No significant difference was detected between distributions based on the nonparametric Mann–Whitney–Wilcoxon test ($P = 0.52$).

Abstracts of published articles that involved one or more clinical trials were reviewed for any mention of original results, regardless of the context. A total of 21 abstracts cited original findings. Of these abstracts, 4 restated principal findings to describe the parent study, and 6 extended the findings to subgroups or to a shorter follow-up period. Of the remaining 11 publications, 1 was a protocol describing simulation of the trial with the use of adaptive designs, 5 involved new statistical approaches, and 5 were reanalyses examining cofactor-by-treatment interactions. Four of these publications were related to reanalyses identified above, whereas the last one described the project more in terms of subgroup analysis than effect modification.

Publications were also examined in terms of published comments. Ascertainment of published comments is described in the Supplementary Appendix. Of the 277 publications that were based on repository clinical trial data, 115 published comments were associated with 57 articles (21%). Of these 115 comments, 12 were author replies, 8 did not have a PubMed-accessible reprint, and 2 were in a language other than English. Full-text reprints were retrieved for the remaining 93 comment articles. None of the comments raised serious concerns regarding misrepresentation of the original study results or serious concerns with the analysis or conclusions; however, 24 (26%) were critical of the article to some degree, such as overstating conclusions, inadequate control of confounders, or questions about the analytic design. One of these commentaries was from the original study investigators who recommended caution in the interpretation of reanalysis results. Commentaries or editorials regarding the clinical significance of the article, future research directions, or how the article may expand the current state of knowledge were found in 58 comment articles (62%). The remainder (11 of 93) consisted of summaries of the article without substantial editorial content.

Discussion

Since inception of the repository, there has been an increasing demand for reuse of clinical trial data. Although there were a variety of reasons for requesting data, the primary use was to examine new hypotheses-generating questions. Kaplan–Meier analysis suggested that approximately 35% of clinical trials were associated with at least one publication within 5 years after release to the general scientific community. In a survey of users of the NHLBI data repository that was conducted by Ross et al., factors that were cited as being burdens to publication included time constraints, lack of funding, lack of statistical support, and a need

for additional data.¹⁵ Citation metrics indicated that resulting publications were, on average, of a similar quality to all publications supported by the NHLBI. Support for released trial data was a non-trivial activity, because approximately one in five requests required some level of assistance.

Requesting clinical trial data for reanalysis or replication of original results was a relatively rare aim among investigators using trial data from the NHLBI repository. Among project descriptions recorded in BioLINCC, only two projects indicated a reanalysis as a primary reason for the data request. This finding is consistent with the survey by Ross et al. of investigators who used data from the NHLBI repository during the 2007–2014 period, in which only seven respondents (4%) indicated that verification of primary results was a principal research objective.¹⁵ That survey of BioLINCC users showed that nearly two thirds of the respondents were new or early-stage investigators (< 10 years of active engagement in clinical research) and that 24% of respondents contacted original investigators to request a collaboration and two thirds of these had the collaboration request accepted, findings that indicate a potential role for repositories in the development of new trialists and epidemiologists.

There is little information regarding what happens once data are released for wide data sharing. In a 2011 commentary based on the National Center for Biotechnology Information Database of Genotypes and Phenotypes (dbGaP), Walker et al. reported on activities during the 2007–2010 period and included 2724 data-access requests from 851 investigators.¹⁶ The principal reason for dbGaP requests involved exploration of new association analysis (39%), followed by methodologic research (26%), replication (18%), and control groups (11%). The dbGaP summary, however, was unable to link requests to subsequent publications. In an evaluation of three open-access platforms with industry-sponsored trial data, Navar et al. found that during the 2013–2015 period, data from approximately 15.5% of more than 3000 available clinical trials had been requested.¹⁷ Of 113 requests with a data-use agreement, 5 indicated validation of the primary end point as a goal of the project, and the lone publication was a validation study in which the results contradicted those of the original publication.¹⁷ An industry-initiated repository reported on the first-year results of open access to approximately 1200 industry-supported trials.¹⁸ Although relatively new, the industry repository reported processing 58 formal requests for data. Strom et al. found that of 177 proposals for data access to the GlaxoSmith-Kline repository (<https://clinicalstudydatarequest.com>), 148 (84%) were for a new study, with few proposals for confirmation of original results (2%)¹⁹ — a utilization pattern similar to that observed for the NHLBI data repository.

There are limitations in our study that should be acknowledged. The NHLBI data repository consists primarily of data from large trials or trials from large clinical networks, and the results of widely sharing these data may not be generalizable to all trials. A crude portfolio analysis was conducted with the use of the NIH Query, View, Report (QVR) system. QVR was searched for funded R01 and U01 (investigator-initiated research project) grants meeting the data-repository submission criteria of new or renewed grants funded on or after January 2006 with direct costs of \$500,000 or more in any given year. Abstracts and aims were reviewed and projects were excluded if there was no interaction with human

participants, the number of participants was less than 500, the project was principally genomic, the project was ancillary to another study, or the project end date was 2018 or later. Of 133 R01s and 81 U01s involving a clinical trial, 38 R01s and 35 U01s were potentially eligible, with only 2 R01s (5%) and 10 U01s (29%) currently in the repository or with known plans to submit. Verification of findings, an oft-cited reason for the wide release of trial data, may be less relevant to the studies in the NHLBI repository owing to established publication histories before repository release. Although considerable effort is expended to discover and validate publications based on repository data, it is likely that some publications were missed. The administrative nature of the data on investigators and data requests introduces inherent limitations in characterizing investigators' experiences with the data repository. Furthermore, because an intent to publish is not a criterion for data access, focusing on publications underestimates the potential value of repository data.

The results of this study of the NHLBI data repository suggest that release of clinical trial data for wide sharing can contribute to the scientific community in multiple ways, including increasing the transparency of findings, examining new hypothesis-generating questions, providing pilot data for grant submissions, testing statistical methods, performing meta-analyses, and developing prediction algorithms. Although the current results provide information on the value of repositories, additional research is certainly warranted. In addition to data on the concerns and attitudes of trialists and study participants, information is needed on a variety of topics related to the wide sharing of data (e.g., the value of a proprietary period for study-generated publications), potential barriers to secondary use of data (e.g., the role of funding), and the value of secondary data analysis in training the next generation of trialists.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Coady SA, Wagner E. Sharing individual level data from observational studies and clinical trials: a perspective from NHLBI. *Trials*. 2013; 14:201. [PubMed: 23837497]
2. Giffen CA, Carroll LE, Adams JT, Brennan SP, Coady SA, Wagner EL. Providing contemporary access to historical bio-specimen collections: development of the NHLBI Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC). *Biopreserv Biobank*. 2015; 13:271–9. [PubMed: 26186276]
3. Biologic Specimens and Data Repository Information Coordinating Center (BioLINCC) home page. Bethesda, MD: National Heart, Lung, and Blood Institute; <https://biolincc.nhlbi.nih.gov/home/>
4. Garrison NA, Sathe NA, Antommaria AH, et al. A systematic literature review of individuals' perspectives on broad consent and data sharing in the United States. *Genet Med*. 2016; 18:663–71. [PubMed: 26583683]
5. Tenopir C, Dalton ED, Allard S, et al. Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS One*. 2015; 10(8):e0134826. [PubMed: 26308551]
6. Federer LM, Lu YL, Joubert DJ, Welsh J, Brandys B. Biomedical data sharing and reuse: attitudes and practices of clinical and scientific research staff. *PLoS One*. 2015; 10(6):e0129506. [PubMed: 26107811]
7. Rathi VK, Strait KM, Gross CP, et al. Predictors of clinical trial data sharing: exploratory analysis of a cross-sectional survey. *Trials*. 2014; 15:384. [PubMed: 25277128]

8. Bell EA, Ohno-Machado L, Grando MA. Sharing my health data: a survey of data sharing preferences of healthy individuals. *AMIA Annu Symp Proc.* 2014; 2014:1699–708. [PubMed: 25954442]
9. Cheah PY, Tangseefa D, Somsaman A, et al. Perceived benefits, harms, and views about how to share data responsibly: a qualitative study of experiences with and attitudes toward data sharing among research staff and community representatives in Thailand. *J Empir Res Hum Res Ethics.* 2015; 10:278–89. [PubMed: 26297749]
10. Saunders PA, Wilhelm EE, Lee S, Merkhofer E, Shoulson I. Data sharing for public health research: a qualitative study of industry and academia. *Commun Med.* 2014; 11:179–87. [PubMed: 26596125]
11. van Panhuis WG, Paul P, Emerson C, et al. A systematic review of barriers to data sharing in public health. *BMC Public Health.* 2014; 14:1144. [PubMed: 25377061]
12. Institute of Medicine Sharing clinical trial data: maximizing benefits, minimizing risks. Washington, DC: National Academies Press; 2015.
13. Taichman DB, Backus J, Baethge C, et al. Sharing clinical trial data — a proposal from the International Committee of Medical Journal Editors. *N Engl J Med.* 2016; 374:384–6. [PubMed: 26786954]
14. Zarin DA, Tse T. Sharing individual participant data (IPD) within the context of the Trial Reporting System (TRS). *PLoS Med.* 2016; 13(1):e1001946. [PubMed: 26784335]
15. Ross JS, Ritchie JD, Finn E, et al. Data sharing through an NIH central database repository: a cross-sectional survey of Bio-LINCC users. *BMJ Open.* 2016; 6(9):e012769.
16. Walker L, Starks H, West KM, Fullerton SM. dbGaP data access requests: a call for greater transparency. *Sci Transl Med.* 2011; 3:113cm34.
17. Navar AM, Pencina MJ, Rymer JA, Louzao DM, Peterson ED. Use of open access platforms for clinical trial data. *JAMA.* 2016; 315:1283–4. [PubMed: 27002452]
18. Strom BL, Buyse M, Hughes J, Knoppers BM. Data sharing, year 1 — access to data from industry-sponsored clinical trials. *N Engl J Med.* 2014; 371:2052–4. [PubMed: 25317745]
19. Strom BL, Buyse ME, Hughes J, Knoppers BM. Data sharing — is the juice worth the squeeze? *N Engl J Med.* 2016; 375:1608–9. [PubMed: 27783903]

Figure 1. Cumulative Requests for Data Sets and Cumulative Number of Available Clinical Trials or Observational Studies

Data are for January 1, 2000, through May 31, 2016.

Figure 2. Fulfilled Data-Access Requests through May 31, 2016, According to Location of Investigator

Shown are the locations of investigators who requested data from at least one clinical trial or observational data only.

Figure 3. Time from Availability of Study Data in the NHLBI Data Repository until Publication of First Article Associated with a Data Request

Data were censored on May 31, 2016. NHLBI denotes National Heart, Lung, and Blood Institute.

Table 1
Characteristics of Investigators and Data Requests That Included at Least One Clinical Trial through May 31, 2016, According to Year Grouping at First Request*

Characteristic	2000–2003	2004–2007	2008–2011	2012–2016 [†]	All Years [‡]
No. of investigators [§]	24	55	96	195	370
Requested multiple data sets — no. (%)	13 (54)	27 (49)	43 (45)	81 (42)	164 (44)
Requested observational data — no. (%)	12 (50)	17 (31)	24 (25)	39 (20)	92 (25)
Type of institution of investigator at first request — no. (%) [¶]					
Academic	18 (75)	37 (67)	83 (86)	149 (76)	287 (78)
Nonprofit	4 (17)	11 (20)	9 (9)	33 (17)	57 (15)
For-profit	2 (8)	7 (13)	4 (4)	13 (7)	26 (7)
Type of data requested at first request — no. (%)					
Cardiovascular					
Epidemiology	7 (29)	9 (16)	13 (14)	20 (10)	49 (13)
Prevention or treatment	14 (58)	31 (56)	43 (45)	99 (51)	187 (51)
Asthma	0	1 (2)	1 (1)	8 (4)	10 (3)
Lung disorders, excluding asthma	3 (12)	12 (22)	29 (30)	39 (20)	83 (22)
Transfusion or transplantation	0	0	2 (2)	5 (3)	7 (2)
Blood disorders	0	0	2 (2)	6 (3)	8 (2)
Adolescents	0	1 (2)	3 (3)	3 (2)	7 (2)
Other	0	1 (2)	3 (3)	15 (8)	19 (5)
Primary reason for first request — no./total no. (%) ^{¶¶}					
Missing data	43/55 (78)	0/96		0/195	
New question	9/55 (16)	77/96 (80)	133/195 (68)	219/303 (72)	
Meta-analysis or pooled studies		0/55	6/96 (6)	15/195 (8)	21/303 (7)
Statistical methods		0/55	3/96 (3)	4/195 (2)	7/303 (2)
Clinical trial analysis		1/55 (2)	6/96 (6)	21/195 (11)	28/303 (9)
Other, such as pilot or simulation study		2/55 (4)	4/96 (4)	22/195 (11)	28/303 (9)
External funding at time of first request — no./total no. (%) ^{**}					
Missing data			40/96 (42)	0/195	
No			38/96 (40)	128/195 (66)	166/251 (66)

Characteristic	2000-2003	2004-2007	2008-2011	2012-2016 [†]	All Years [‡]
Yes					
Source of external funding — no./total no. (%)					
NIH Extramural Award			13/18 (72)	27/67 (40)	40/85 (47)
NIH Intramural Award			0/18	3/67 (4)	3/85 (4)
Non-NIH federal funding			2/18 (11)	6/67 (9)	8/85 (9)
Private foundation			0/18	10/67 (15)	10/85 (12)
Non-U.S. funding			2/18 (11)	9/67 (13)	11/85 (13)
Other external funding			1/18 (6)	12/67 (18)	13/85 (15)
Country of residence — no. (%)					
United States or Canada	23 (96)	52 (95)	86 (90)	164 (84)	325 (88)
United Kingdom	0	2 (4)	2 (2)	15 (8)	19 (5)
Other country in Europe	1 (4)	1 (2)	4 (4)	9 (5)	15 (4)
Asia	0	0	1 (1)	3 (2)	4 (1)
Australia or New Zealand	0	0	3 (3)	2 (1)	5 (1)
Other	0	0	0	2 (1)	2 (1)
Request was associated with 1 publication involving trial data — no. (%)	11 (46)	20 (36)	47 (49)	33 (17)	111 (30)

* NIH denotes National Institutes of Health.

[†] Data are through May 31, 2016.

[‡] Investigators with missing values were not included in the calculation of percentages.

[§] Shown is the number of investigators initiating their first data-repository request.

[¶] Academic institutions were defined as clearly academic (university or college) or by the investigator's having an e-mail address ending in ".edu." Nonprofit institutions were defined by an e-mail address ending in ".org." For-profit institutions were defined as clearly commercial entities or by an e-mail address ending in ".com."

^{||} The primary reason for the request was inferred from project titles beginning in August 2007.

^{**} Investigators' reporting of external funding began in October 2009.

Characteristics of Clinical Trials with Data in the NHLBI Data Repository, According to Year Grouping of Study Release Date, through May 31, 2016*

Table 2

Characteristic	2000–2003	2004–2007	2008–2011	2012–2016 [†]	All Years
No. of trials	16	27	20	37	100
No. of participants					
Median	2519	2708	602	551	924
Mean	4646±6937	5507±12,906	3186±9416	1688±3380	3492±8646
Year that study started — no. (%)					
1971–1990	9 (56)	11 (41)	2 (10)	1 (3)	23 (23)
1991–2000	7 (44)	16 (59)	11 (55)	6 (16)	40 (40)
2001–2005	0	0	7 (35)	8 (22)	15 (15)
2006–2010	0	0	0	22 (59)	22 (22)
Major medical or population group — no. (%)					
Adolescents	0	1 (4)	1 (5)	0	2 (2)
Asthma	2 (12)	3 (11)	6 (30)	3 (8)	14 (14)
Blood disorders	0	0	1 (5)	3 (8)	4 (4)
Cardiovascular disorders	9 (56)	19 (70)	7 (35)	16 (43)	51 (51)
Lung disorders, excluding asthma	4 (25)	3 (11)	2 (10)	8 (22)	17 (17)
Transfusion or transplantation	0	1 (4)	3 (15)	2 (5)	6 (6)
Other	1 (6)	0	0	5 (14)	6 (6)
Intervention type — no. (%)					
Drug	5 (31)	12 (44)	10 (50)	16 (43)	43 (43)
Behavioral	1 (6)	6 (22)	5 (25)	5 (14)	17 (17)
Procedure	3 (19)	5 (19)	4 (20)	2 (5)	14 (14)
Multiple	7 (44)	3 (11)	1 (5)	9 (24)	20 (20)
Other, such as radiation or biologic	0	1 (4)	0	5 (14)	6 (6)
Purpose — no. (%)					
Diagnostic	0	1 (4)	0	1 (3)	2 (2)
Prevention	4 (25)	13 (48)	9 (45)	7 (19)	33 (33)
Treatment	12 (75)	13 (48)	11 (55)	29 (78)	65 (65)

Characteristic	2000–2003	2004–2007	2008–2011	2012–2016 [‡]	All Years
At least 1 publication associated with request — no. (%)	9 (56)	19 (70)	12 (60)	7 (19)	47 (47)

* Plus–minus values are means $\pm$ SD. NHLBI denotes National Heart, Lung, and Blood Institute.

[‡] Data are through May 31, 2016.