BMJ Open

Data Sharing through a NIH Central Database Repository: A cross-sectional survey of BioLINCC users

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012769
Article Type:	Research
Date Submitted by the Author:	22-May-2016
Complete List of Authors:	Ross, Joseph; Yale University School of Medicine, Internal Medicine Ritchie, Jessica; Yale-New Haven Hospital, Center for Outcomes Research and Evaluation Finn, Emily; Yale-New Haven Hospital, Center for Outcomes Research and Evaluation Desai, Nihar; Yale-New Haven Hospital, Center for Outcomes Research and Evaluation Lehman, Richard; Oxford University, Primary Health Care Krumholz, Harlan; Yale University, Medicine Gross, Cary; Yale School of Medicine, Section of General Internal Medicine
Primary Subject Heading:	Medical publishing and peer review
Secondary Subject Heading:	Health policy, Health services research, Evidence based practice, Ethics
Keywords:	Clinical trials < THERAPEUTICS, STATISTICS & RESEARCH METHODS, PUBLIC HEALTH

Data Sharing through a NIH Central Database Repository:

A cross-sectional survey of BioLINCC users

Joseph S Ross, MD, MHS, *associate professor of medicine*[1,2,3,4]; Jessica D Ritchie, MPH, PMP, *research coordinator*[1]; Emily Finn, MPH, *research associate*[2]; Nihar R Desai, MD, *assistant professor of medicine*[1,5]; Richard L Lehman, MD, *senior advisory fellow*[1,6]; Harlan M Krumholz, MD, SM, *Harold H Hines Jr, professor of medicine and epidemiology and public health*[1,3,4,5]; Cary P Gross, MD, *professor of medicine*[2,3,7].

[1] Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, CT 06510, USA;

[2] Section of General Internal Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT 06510, USA;

[3] Robert Wood Johnson Foundation Clinical Scholars Program, Department of Internal Medicine, Yale School of Medicine, New Haven, CT 06510, USA;

[4] Department of Health Policy and Management, Yale School of Public Health, New Haven, CT 06520, USA;

[5] Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT 06510, USA;

[6] UK Cochrane Center, Oxford, England;

[7] Cancer Outcomes, Public Policy, and Effectiveness Research Center, Yale Cancer Center, New Haven, CT 06510, USA.

Note: Ms. Finn was affiliated with the Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, during the time the work was conducted.

Correspondence: Joseph S. Ross, MD, MHS, Section of General Internal Medicine, Yale University School of Medicine, P.O. Box 208093, New Haven, CT 0520, Phone: (203) 785-2987, Fax: (203) 737-3306, email: joseph.ross@yale.edu.

Word count (text only): 3802 **Tables:** 2 **References:** 20

Abstract word count: 324 **Figures:** 3

Key words: Clinical trials; Data sharing; Open science.

ABSTRACT

Objective

To characterize experiences with using clinical research data shared through the National Institute of Health's Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC) clinical research data repository, along with data recipients' perceptions of the value, importance, and challenges with using BioLINCC data.

Design and Setting

Cross-sectional web-based survey.

Participants

All investigators who requested and received access to clinical research data from BioLINCC between 2007 and 2014.

Main Outcome Measures

Reasons for BioLINCC data request, research project plans, interactions with original study investigators, BioLINCC experience, and other project details.

Results

There were 536 investigators who requested and received access to clinical research data from BioLINCC between 2007 and 2014. Of 441 potential respondents, 195 completed the survey (response rate=44%); 89% (n=174) requested data for an independent study, 17% (n=33) for pilot/preliminary analysis. Commonly cited reasons for requesting data through BioLINCC were feasibility of collecting data of similar size and scope (n=122) and insufficient financial resources for primary data collection (n=76). For 95% of respondents (n=186), a primary research objective was to complete new research, as opposed to replicate prior analyses. Prior to requesting data from BioLINCC, 18% (n=36) of respondents had contacted the original study investigators to obtain data, whereas 24% (n=47) had done so to request collaboration. Nearly all (n=176; 90%) respondents found the data to be suitable for their proposed

project; among those who found the data unsuitable (n=19; 10%), cited reasons were data too complicated to use (n=5) and data poorly organized (n=5). Half (n=98) of respondents had completed their proposed projects, of which 67% (n=66) have been published.

Conclusions

Investigators were primarily using clinical research data from BioLINCC for independent research, making use of data that would otherwise have not been feasible to collect.

Article Summary: Strengths and Limitations of this Study

- Data sharing policies are increasingly promoted and being adopted by research funders to improve access to clinical trial data to inform evidence-based practice. The National Institute of Health's Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC) has been actively sharing data from its clinical research data repository for more than 10 years.
- In the first survey of the experiences of investigators who have requested and been approved to use data from BioLINCC, we found that users were primarily focused on conducting independent research studies, making use of data that would otherwise have not been feasible to collect, because of both insufficient time and resources.
- We also found that shared data from BioLINCC could be used to successfully pursue clinical research; 90% of BioLINCC users found the data to be suitable, half had completed their research projects thus far, and two-thirds had published their findings.
- Our study of user experiences with BioLINCC offer important insights for newly initiated and on-going clinical trial data sharing efforts and illustrate the potential and value of data sharing for the broader scientific field, as well as the challenges that remain to be overcome. low response rate, and may be affected by recall bias and social desirability bias, perhaps suggesting that our findings overestimate the perceived value of BioLINCC data and its usability for the broader scientific community.

Over the past 5 years, several major research funders, including the U.S. National Institutes of Health (NIH), the U.S. Patient-Centered Outcomes Research Institute, the U.K. Medical Research Council, and the Bill and Melinda Gates Foundation, as well as private industry,[1] have adopted policies supporting or mandating clinical research data sharing. In January 2015, the Institute of Medicine of the U.S. National Academies further supported these efforts with its report, "Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risks," recommending that stakeholders foster a culture in which data sharing is the expected norm and commit to responsible strategies aimed at maximizing benefits, minimizing risks, and overcoming challenges of sharing clinical trial data.[2] In January 2016, the International Committee of Medical Journal Editors issued a proposal to require authors to share with others the deidentified individual-patient data underlying the results presented in the article no later than 6 months after publication as a condition of consideration for publication of a clinical trial report in our member journals.[3]

In response to these new policies and proposals, funded investigators will increasingly be asked to prepare and make collected data available to other investigators with whom they are not collaborating so that the second can pursue independent research. To support these efforts and inform developing policies, a number of prior studies have examined the willingness of clinical trial investigators to share clinical research data, generally finding broad support, and characterized anticipated challenges to and concerns with data sharing.[4-11] However, few studies have focused on the investigators who have actually received deidentified individual-patient data from a centralized data sharing platform, in order to understand their perspectives regarding challenges encountered with requesting and using the data, and disseminating findings.

While most of these data sharing efforts have been relatively newly established, the U.S. National Heart, Lung, and Blood Institute (NHLBI) of the NIH established a formal data repository in 2000, now managed by the Biologic Specimen and Data Repository Information Coordinating Center

(BioLINCC), to facilitate access to, maximize the scientific value of, and promote the availability and use of the biorepository, data repository and other NHLBI-funded population-based biospecimen and data resources.[12 13] As BioLINCC has been actively sharing data for more than a decade and currently receives over 100 requests for clinical trial and other prospective cohort clinical data per year (ref: personal communication, Sean Coady, NHLBI data repository manager), there is an opportunity to learn from data users' experiences to inform clinical data sharing efforts. Accordingly, we surveyed all investigators who requested and received access to clinical research data from BioLINCC between 2007 and 2014. We specifically sought to understand their experiences with clinical research data sharing and status of their research project, as well as perceptions of the value, importance, and challenges of accessing data through BioLINCC.

METHODS

Study Sample and Design

We conducted a cross-sectional survey from May to August 2015 of all investigators who requested and received access to clinical research data from BioLINCC between 2007 and 2014. This time period was chosen to ensure a contemporaneous sample of investigators whose contact information was less likely to have changed over ensuing years. In accordance with NIH policy, BioLINCC provided our study team with a list of investigators who had requested and received access using a *public e-mail address*; contact information was available for the lead investigator who was responsible for the BioLINCC request, not each member of the study team. For investigators who had requested and received access using a *private e-mail address*, BioLINCC first sent an opt-in/opt-out e-mail in May 2015 asking if they would be willing to participate in the survey (**Appendix**). Non-respondents were sent two follow-up requests by e-mail; those that did not respond by the end of the third week were considered

to have opted-out. BioLINCC subsequently provided our study team with a list of those investigators who opted-in.

In addition to contact information, BioLINCC provided our study team with information on the following for all investigators who had requested and received access to clinical research data: lead investigator location, affiliation with an academic institution or for-profit organization, and total number of requests ever submitted to BioLINCC, as well as the request year, the number of data sets requested, and self-reported availability of external funding to support the research project using the requested data.

In May 2015, the Yale team sent all potential survey respondents an initial e-mail to describe the purpose of the study, request their participation, and provide a link to the survey; three follow-up requests were sent by e-mail over the course of June 2015. Non-respondents were contacted by telephone to solicit their participation up to twice per week, but no more than once per day, until one contact was made. In July 2015, Internet searches to update contact information for non-respondents were conducted. For all non-respondents for whom updated contact information was identified, the initial survey email was sent, followed by three follow-up requests.

Invitations to participate did not reference a specific hypothesis of the study, but stated that investigator participation would further the understanding of investigators' experience with BioLINCC and inform future clinical trial data sharing efforts (**Appendix**). Participation was voluntary and included an opportunity to win one of five $100 gift certificates for Amazon. All internet-based responses were collected using a Web-based survey platform (Qualtrics Labs, Provo, UT). Approval from the Yale University School of Medicine Human Research Protection Program was obtained prior to study conduct and consent was considered to be implied when participants completed the online survey.

Survey Instrument Development

The design of our 50-item survey instrument was informed by previously published surveys,[4 5] a review of the literature on clinical trial data sharing, and discussion with multiple experts and stakeholders, including representatives from NHLBI and academic investigators. Experts recommended survey topics that they considered to be compelling for the field of data sharing and re-use of data. The survey was pre-tested with six medical students and staff at the Yale-New Haven Hospital Center for Outcomes Research and Evaluation (New Haven, CT) and modified iteratively to improve clarity, face validity, and content validity. Adaptive questioning was used to decrease response burden. Items were presented in multiple response, Likert scale, and open-ended formats; many of the multiple response questions enabled respondents to select multiple answers. The complete instrument is provided within the **Appendix**.

Survey Domains

Reasons for Data Request and Planned Research Project

We used multiple response and yes/no questions to assess investigators' primary research purpose and reasons for requesting data from BioLINCC. Multiple response questions were also used to determine the primary research objective, funding used to support the project, and other details of the planned research project.

Interactions with Original Study Investigators

We used yes/no questions to determine whether original study investigators were contacted prior to or after requesting data through BioLINCC to obtain the data or to collaborate. These were followed by multiple response questions to determine why collaborations were sought, whether the requests for data or collaboration were approved, and reasons for not approving.

BioLINCC Experience

Multiple response, yes/no, and Likert-type questions were used to obtain information regarding investigator's experience using BioLINCC, including whether the data were suitable and useful for their project.

Project Details

We used multiple response and yes/no questions to characterize the completion stage of investigators' projects. For those that did not complete their project, multiple response and yes/no questions were used to ascertain reasons why the project was incomplete. For those with completed projects, we used multiple response and yes/no questions to determine whether the final project differed from the pre-specified project as well as to obtain publication information. Multiple choice and multiple response questions were used to identify any funding sources and whether using the data from BioLINCC aided in any future grant applications.

Requestor Demographics

Respondents were asked to characterize their primary employer and career status using multiple choice questions, including whether they had ever been closely involved (as Principal or Co-Investigator) in the conduct of a randomized controlled trial and/or ever deposited clinical trial data in the BioLINCC repository. Respondent sociodemographic characteristics, including age, gender, and ethnicity, were also collected.

Patient Involvement

Patients were not involved in the design or conduct of this study. Results will be directly disseminated via email to all individuals invited to participate in the survey upon publication.

Statistical Analysis

To compare characteristics of survey respondents and non-respondents, we used two-sided Chi-square tests and Fisher Exact tests when appropriate with a type 1 error level of 0.05. Next, we conducted descriptive analyses of the reasons for requesting data from BioLINCC, prior interactions with original trial investigators, experience using BioLINCC, and project details, as well as respondent demographic characteristics. Data were analyzed using JMP Pro Version 11.2.0 (SAS Institute Inc., Cary, NC).

RESULTS

There were 536 investigators who requested and received access to clinical research data from BioLINCC between 2007 and 2014 (**Figure 1**). Investigators for which a public e-mail address was not available were sent an opt-in/opt-out letter (n=74); 23 opted in, 3 opted out, 7 could not be reached, and 41 were not responsive. Survey participation requests were thus sent to 485 eligible respondents, 44 of whom were subsequently excluded due to the following reasons: invalid contact information (n=31), the investigator had no recollection of requesting the data (n=5), or the data had been requested by someone other than the investigator (n=8). Of the remaining 441 respondents, 195 completed the survey, yielding a survey response rate of 44.2%. However, of the 536 total investigators who requested and received access to clinical research data from BioLINCC, 195 completed the survey (response rate of 36.3%).

Survey respondents did not differ from non-respondents with respect to investigator location, affiliation with an academic institution or for-profit organization, and total number of requests ever submitted to BioLINCC, as well as the number of data sets requested (P values ≥ 0.10; **Table 1**).

However, respondents were more likely than non-respondents to have requested data more recently (P=0.004) and to have self-reported external funding to support the research project (P=0.009).

Half of survey respondents were between the ages of 35 and 49 years old (n=97; 50%), while 59% were male (n=116), 68% were white (n=133), and 90% identified as not Hispanic/Latino (n=175; **Table 2**). The vast majority of respondents were primarily employed by an academic institution (n=165; 85%) and 78% (n=152) have been engaged in clinical research for at least three years. While 42% (n=82) had been closely involved in the conduct of a randomized controlled trial, only 3% (n=5) had ever deposited data in the BioLINCC repository.

Reasons for Data Request

Overall, respondents' motivations for requesting data from BioLINCC were largely focused on using the data to conduct and disseminate new research studies, as 89% (n=174) indicated that data were requested for an independent study, 17% (n=33) to use the data for pilot/preliminary analysis. For 63% (n=122) of respondents, the decision to request data through BioLINCC was influenced by the belief that collecting data of similar size and scope was not feasible, while insufficient financial resources for primary data collection (n=76; 39%), individual participant-level data being unavailable elsewhere (n=71; 36%), and insufficient time for primary data collection (n=64; 33%) were also commonly cited reasons for requesting data through BioLINCC (**Figure 2**).

Planned Research Project

Respondents largely (n=149; 76%) planned research projects that used the requested BioLINCC data as a standalone data source for at least one project, while 35% (n=69) planned to combine the data with other data sources; of these, 32% (n=22) planned to conduct a meta-analysis. Nearly all respondents (n=186; 95%) indicated that at least one of their primary research objectives was to

complete new research, whereas only 7 (4%) had a primary research objective solely to replicate prior analyses. Of those pursuing new research, 56% (n=104) planned to leverage the data for a research question unrelated to the original research design, while 40% (n=74) planned to examine subgroup populations and 32% (n=60) planned to examine secondary endpoints.

Only 13% (n=26) of respondents indicated that the focus of their research was a medical product or intervention; of these, 73% (n=19) planned analyses to examine product / intervention efficacy, 54% (n=14) safety. Finally, 58% (n=114) of respondents had funding to support the research project, most commonly from the NIH (n=44; 23%), whereas 43% (n=84) primarily self-funded the research project.

Interactions with Original Study Investigators

Fewer than one in five (n=36; 18%) respondents indicated that they had contacted the original study investigators to obtain data prior to requesting the data from BioLINCC; among these, 44% (n=16) reported that the original study investigator approved their request and these investigators most commonly requested access to the data from BioLINCC anyway because the process to access data was more straightforward through BioLINCC (n=11). Among the 20 (56%) respondents who indicated that the original study investigator denied their request, the most common response given by the original investigator was to direct the respondent to BioLINCC (n=11; 55%).

Nearly one-quarter of respondents (n=47; 24%) indicated that they contacted the original study investigator to request collaboration, most commonly because of an interest in working with the original study investigators (n=23) and need for additional content expertise due to study design complexity (n=20). Of the respondents who requested collaboration, two-thirds (n=31; 66%) indicated that the request was accepted.

Data Repository Experience

Nearly all respondents indicated satisfaction with the data available through BioLINCC and that they were suitable for their originally proposed project (n=176; 90%). Among the 19 (10%) respondents who indicated that the data were not suitable, the two most commonly cited reasons were that the data were too complicated to use, preventing them from determining whether the data were suitable (n=5); and that the data were poorly organized, preventing adequate preparation for analysis (n=5).

Research Project Details

Half of all respondents (n=98; 50%) reported that their projects have been completed, of which 67% (n=66) have been published. Of those who have completed their research, 48% (n=47) indicated that no substantive concerns were raised about the use of data from BioLINCC during the peer-review process, while 8% indicated that concerns were raised about research methodology and analysis (n=8), 7% about the original study design that the investigator could not address (n=7), and 6% about their research project design that they could not address without additional data (n=6).

Of the 50% of respondents who have not yet completed their proposed projects (n=97), 84% (n=81) explained that they planned to complete their project; 65% (n=63) indicated that their project is in analysis/manuscript draft phase, while 28% (n=27) explained that they have thus far been too busy with other responsibilities to complete the research project using the data from BioLINCC and 13% (n=13) reported that lack of funding to support the project was a problem (**Figure 3**). Sixteen investigators explained that they did not intend to complete their project, most often because the age of the data made the project now less relevant or because of data issues, such as missing values for the variable of interest.

Of the 179 respondents who already completed or planned to complete their proposed project, 54% (n=96) reported that there would be one research project resulting from their single request for data from BioLINCC, 23% (n=42) reported two, and 23% (n=41) reported three or more. In addition, 15%

(n=27) of respondents who have completed or planned to complete their project indicated that their completed/anticipated final project differed from their pre-specified project; the most commonly modified aspects were the statistical analysis plan (n=18) and the selection of the main independent variables (n=12).

DISCUSSION

In this survey of investigators who had requested and received access to clinical research data from BioLINCC between 2007 and 2014, the vast majority had requested the data in order to conduct independent research projects, primarily because collecting data of similar size and scope was not feasible, due to both insufficient time and resources. Half of the investigators had completed their research projects thus far, two-thirds of which published their findings, and among those investigators whose projects were incomplete, two-thirds were actively engaged in analysis or manuscript preparation. These findings offer important insights for newly initiated and on-going clinical trial data sharing efforts and illustrate the potential and value of data sharing for the broader scientific field, as well as the challenges that remain to be overcome.

First, the BioLINCC experience suggests that when clinical research data are made available to investigators, there is likely to be interest in using the data for independent research projects. There are currently 654 publications associated with the data repository available through BioLINCC.[14] This large number of publications suggests that these data are being used by investigators, better maximizing the NHLBI investment in and scientific value of clinical research data. Many investigators responding to our survey noted that collecting data of similar size and scope was not feasible, or that they had insufficient financial resources or time for primary data collection, justifying the need to request data from BioLINCC for their research.

Second, the BioLINCC experience suggests that clinical data can be collected by one set of investigators and made available to another set of investigators who, for the most part, can use it to successfully pursue an independent research project. While some surveyed investigators noted challenges in using the data made available through BioLINCC, 90% found the data to be suitable for their originally proposed project, even without input from the original research team. Few reported that the data were too complicated to use, preventing them from determining whether the data were suitable, or that the data were poorly organized.

Finally, the research enterprise is not optimally efficient, and the BioLINCC experience reflects this short-coming. In aggregate, more than 100 research projects were completed as a result of respondent investigators using data made available through BioLINCC. However, despite all investigators having received data from BioLINCC at no cost, only half of investigators who had received data had completed their research projects thus far. While many more continue to work on their projects and intend to complete their work, the investment by NHLBI to make these data available should be matched by the effort of investigators to ensure that the projects are completed. Moreover, even among completed projects, only two-thirds were published. Mechanisms should be established to ensure that results from research made possible through data sharing are publicly disseminated, either through publication or through a results reporting initiative similar to ClinicalTrials.gov.

For the potential and value of data sharing to be fully realized, more needs to be accomplished. Part of the success of BioLINCC may be attributed to the NHLBI policy that supported studies with direct costs equal to or greater than $500K in any 1 year and identified as being of high programmatic interest, along with co-operative agreements with 500 or more participants are required to submit data as part of the grant award.[13] This policy establishes clear expectations for data sharing, so that data can be properly organized and de-identified and supportive documentation and materials prepared in anticipation of submitting data to BioLINCC. However, it's not clear whether this policy allows

researchers to budget resources for this work. Currently, the NIH is seeking ways to broaden data sharing efforts across its institutes;[15] to enhance the likelihood of success of data sharing efforts, it should be clarified whether NIH-granted independent research funds can be used to prepare collected data for sharing through initiatives such as BioLINCC.

Similarly, financial support for investigators to use clinical research data that are being shared and made available would enhance efforts. Forty percent of investigators using data from BioLINCC had self-funded their research efforts, while an eighth were relying on funding from the NIH. However, among surveyed investigators who had not yet completed their proposed projects, lack of funding to support the project was a commonly cited problem. Without financial support, efforts to share data are likely to fail to achieve their potential,[2] even despite the strong policies and proposals in favor of data sharing from other research funders, the Institute of Medicine, and the International Committee of Medical Journal Editors.

There are important limitations of our study to consider. First, only 44% of potentially eligible respondents completed our survey, perhaps suggesting that our findings overestimate the perceived value of BioLINCC data and its usability for the broader scientific community. Individuals who chose not to respond to our survey may have found the data to be more problematic and less useful than those who responded. Furthermore, even among respondents, our findings may have been biased by recall bias, including an inability to remember using the data made available by BioLINCC, and social desirability,[16 17] as respondents may have been less likely to self-report experiences and project completion plans that may be negatively perceived by others. In addition, there were a few observed differences between survey respondents and non-respondents. Because we would expect that investigators who made more recent requests and who had secured external funding to support the research project would be more likely to remain enthusiastic about the project and to complete it, our findings may be biased toward higher project completion rates. However, our response rate compares

favorably with other surveys of physicians and investigators,[4 18-20] perhaps reflecting that we used several mechanisms to prospectively improve response rates, including a web-based survey platform for ease of completion, we employed several reminder contacts, including three e-mails and at least one telephone contact and we offered financial incentives for participation .

Second, our study was limited to investigators who had received data from BioLINCC and our findings may not be applicable to the experience of investigators obtaining data from other repositories. Third, some information of interest was not asked in order to reduce survey response burden, including questions asking about the time and effort invested to manage and analyze the data from BioLINCC and the impact of the publications resulting from the research project. Finally, our study made no attempt to judge the impact of the research that was able to be completed because of the clinical research data made available through BioLINCC. Other efforts should consider whether the investment being made by NIH and NHLBI in data sharing is justified by the information and knowledge being generated for medical science and society.

In conclusion, we found that the vast majority of investigators who had requested and received access to clinical research data from BioLINCC between 2007 and 2014 had either succeeded in completing their research project or reported being actively involved data analysis or manuscript preparation. In aggregate, more than 100 research projects were completed as a result of respondent investigators using data made available through BioLINCC. Experience with BioLINCC illustrates the potential of data sharing for the broader scientific field and the importance of funding these efforts, particularly when collecting data of similar size and scope is not feasible for many investigators.

Acknowledgements

The authors would like to acknowledge Mr. Sean Coady of the National Heart, Lung, and Blood Institute (NHLBI) of the National Institutes of Health (NIH), who provided feedback during survey development and assistance with contact information for potential survey respondents, as well as Ms. Tiffany Chang, MPH, Ms. Julia Eichenfield, and Mr. Daniel Shaw, who provided background research and assistance with survey distribution during the course of student employment / voluntary summer research experience at the Yale-New Haven Hospital Center for Outcomes Research and Evaluation (CORE).

Contributors: Study concept and design (Ross, Krumholz, Gross); Acquisition of data (Ritchie, Finn); Analysis and interpretation of data (all authors); Drafting of manuscript (Ross, Ritchie, Finn); Critical revision of manuscript (all authors); Statistical analysis (Ross, Finn); Study supervision (Ross).

Funding/support and role of the sponsor: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. The authors assume full responsibility for the accuracy and completeness of the ideas presented.

Competing Interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and all authors declare (currently or formerly) receiving support through Yale University from Medtronic, Inc. and Johnson and Johnson to develop methods of clinical trial data sharing and from the Blue Cross Blue Shield Association (BCBSA) to better understand medical technology evidence generation. Drs. Krumholz and Ross receive support through Yale University from the Centers of Medicare and Medicaid Services (CMS) to develop and maintain performance measures that are used for public reporting and from the Food and Drug Administration (FDA) to develop methods for post-market surveillance of medical devices. Dr. Krumholz chairs a cardiac scientific advisory board for UnitedHealth. No other disclosures were reported.

Ethical Approval

Ethics approval from the Yale University School of Medicine Human Research Protection Program was obtained prior to study conduct and consent was considered to be implied when participants completed the online survey.

Data sharing

Requests for statistical code and the dataset can be made to the corresponding author at joseph.ross@yale.edu. The dataset will be made available via a publicly accessible repository on publication, at the Dryad Digital Repository (datadryad.org).

Transparency

The lead author (JSR) affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

License

REFERENCES

1. Pharmaceutical Research and Manufacturers of America (PhRMA) and European Federation of Pharmaceutical Industries and Associations (EFPIA). Principles for Responsible Clinical Trial Data Sharing. Our Committment to Patients and Researchers. Secondary Principles for Responsible Clinical Trial Data Sharing. Our Committment to Patients and Researchers. July 2013. http://phrma.org/sites/default/files/pdf/PhRMAPrinciplesForResponsibleClinicalTrialDataSharing.pdf.
2. Institute of Medicine of the National Academies of Science. *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risks*. Washington, DC: National Academies Press, 2015.
3. Taichman DB, Backus J, Baethge C, et al. Sharing Clinical Trial Data: A Proposal From the International Committee of Medical Journal Editors. Annals of internal medicine 2016.
4. Rathi V, Dzara K, Gross CP, et al. Sharing of clinical trial data among trialists: a cross sectional survey. BMJ 2012;**345**:e7570.
5. 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.
6. Savage CJ, Vickers AJ. Empirical study of data sharing by authors publishing in PLoS journals. PloS one 2009;**4**(9):e7078.
7. Tenopir C, Allard S, Douglass K, et al. Data sharing by scientists: practices and perceptions. PloS one 2011;**6**(6):e21101.
8. 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.
9. Oushy MH, Palacios R, Holden AE, et al. To Share or Not to Share? A Survey of Biomedical Researchers in the U.S. Southwest, an Ethnically Diverse Region. PloS one 2015;**10**(9):e0138239.

10. Tudur Smith C, Dwan K, Altman DG, et al. Sharing individual participant data from clinical trials: an opinion survey regarding the establishment of a central repository. PloS one 2014;**9**(5):e97886.

11. Hopkins C, Sydes M, Murray G, et al. UK publicly funded Clinical Trials Units supported a controlled access approach to share individual participant data but highlighted concerns. Journal of clinical epidemiology 2016;**70**:17-25.

12. Giffen CA, Carroll LE, Adams JT, et al. Providing Contemporary Access to Historical Biospecimen Collections: Development of the NHLBI Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC). Biopreservation and biobanking 2015;**13**(4):271-9.

13. Coady SA, Wagner E. Sharing individual level data from observational studies and clinical trials: a perspective from NHLBI. Trials 2013;**14**:201.

14. U.S. National Institutes of Health, National Heart L, and Blood Institute (NHLBI),, Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC). Publications associated with available BioLINCC resources. Secondary Publications associated with available BioLINCC resources. https://biolincc.nhlbi.nih.gov/publications/.

15. Hudson KL, Collins FS. Sharing and reporting the results of clinical trials. JAMA : the journal of the American Medical Association 2015;**313**(4):355-6.

16. Fowler FJ. *Survey Research Methods*. Newbury Park, CA: Sage, 1993.

17. Sudman S, Bradburn NM. *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco, CA: Jossey-Bass, 1982.

18. Asch DA, Jedrziewski MK, Christakis NA. Response rates to mail surveys published in medical journals. Journal of clinical epidemiology 1997;**50**(10):1129-36.

19. Keyhani S, Federman A. Doctors on coverage--physicians' views on a new public insurance option and Medicare expansion. The New England journal of medicine 2009;**361**(14):e24.

20. Shanafelt TD, Boone S, Tan L, et al. Burnout and Satisfaction With Work-Life Balance Among US Physicians Relative to the General US Population. Archives of internal medicine 2012:1-9.

FIGURE LEGENDS

Figure 1: Inclusion flow chart used to identify potential survey respondents: investigators who had requested and received access to clinical research data from BioLINCC between 2007 and 2014.

Figure 2: Factors influencing decision to request clinical research data through BioLINCC between 2007 and 2014 (n=195).

Figure 3. Flow chart showing completion rates of research projects using clinical research data requested from BioLINCC between 2007 and 2014.

Table 1. Characteristics of survey respondents and non-respondents.

	Respondents, No. (%) (n=195)	Non-respondents, No. (%) (n=246)	P value
Investigator based in the US?			
Yes	163 (84)	211 (86)	0.53
No	32 (16)	35 (14)	
Investigator based at academic institution?			
Yes	149 (76)	196 (80)	0.41
No	46 (24)	50 (20)	
Investigator based at for-profit institution?			
Yes	5 (3)	4 (2)	0.49
No	190 (97)	242 (98)	
Investigator's total submitted requests to BioLINCC (ever), No.			
1	120 (62)	169 (69)	0.12
> 1 (includes renewals)	75 (38)	77 (31)	
Data sets requested, No.			
1	152 (78)	171 (70)	0.10
2-4	31 (16)	58 (24)	
5-9	7 (4)	14 (6)	

10+	5 (3)	3 (1)	
Request year			
2006	0 (0)	1 (<1)	
2007	13 (7)	17 (7)	
2008	4 (2)	16 (7)	
2009	9 (5)	23 (9)	
2010	6 (3)	17 (7)	0.004
2011	12 (6)	26 (11)	
2012	43 (22)	55 (22)	
2013	47 (24)	39 (16)	
2014	61 (31)	52 (21)	
External funding to support the research project?			
Yes	74 (38)	77 (31)	
No	97 (50)	111 (45)	0.009
Unknown	24 (12)	58 (24)	

Table 2. Sociodemographic and professional characteristics of survey respondents (n=195).

Characteristic	No (%) of respondents
Age	
34 years or younger	29 (15)
35-49 years	97 (50)
50-64 years	47 (24)
65 years or older	14 (7)
Prefer not to answer	8 (4)
Gender	
Male	116 (59)
Female	74 (38)
Prefer not to answer	5 (3)
Race	
White	133 (68)
Asian	35 (18)
Black or African American	10 (5)
Other	3 (2)
Prefer not to answer	14 (7)
Ethnicity	
Hispanic or Latino	8 (4)
Not Hispanic or Latino	175 (90)
Prefer not to answer	12 (6)
Primary employer	

Academic Institution	165 (85)
Non-Profit Organization	14 (7)
Government	8 (4)
Private Industry	4 (2)
Other	4 (2)
Career stage	
In training (< 3 years of active engagement in clinical research, still receiving formative training in research methods)	43 (22)
Early stage career (3-10 years of active engagement in clinical research)	83 (43)
Established in the field (> 10 years of active engagement in clinical research)	69 (35)
Ever been closely involved (as PI or Co-PI) in the conduct of a randomized controlled trial?	
Yes	82 (42)
Ever deposited clinical trial data in the BioLINCC repository?	
Yes	5 (3)

Figure 1: Inclusion flow chart used to identify potential survey respondents: investigators who had requested and received access to clinical research data from BioLINCC between 2007 and 2014.

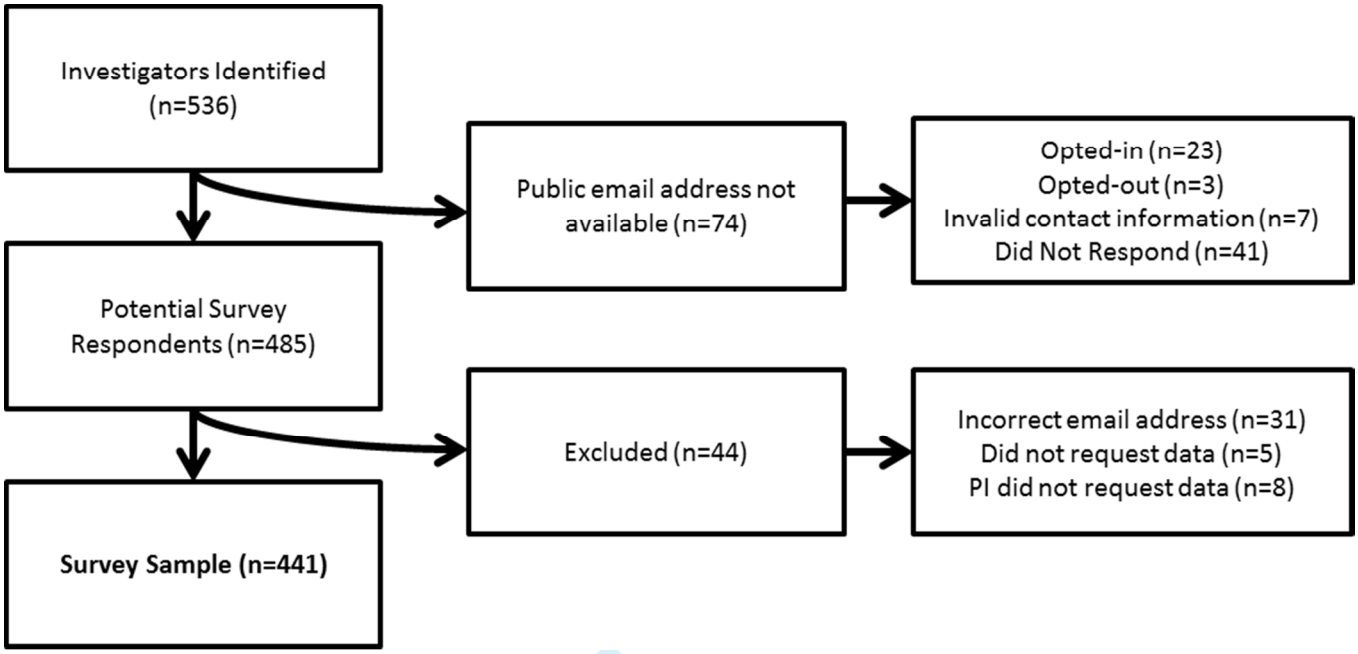

Figure 2: Factors influencing decision to request clinical research data through BioLINCC between 2007 and 2014 (n=195).

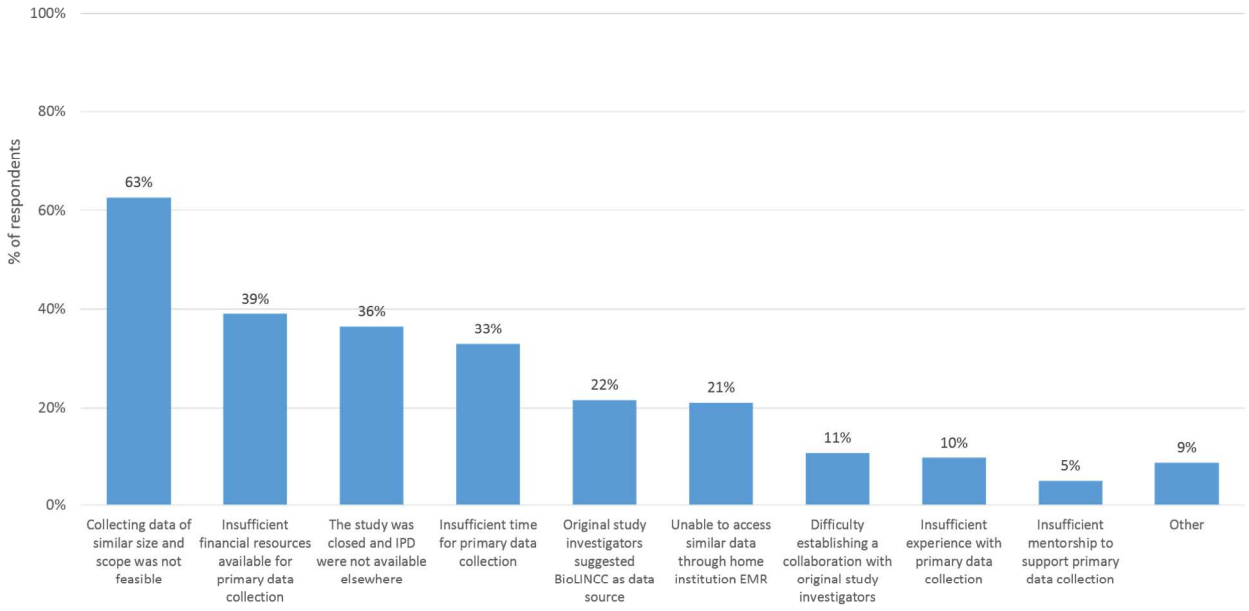

Fig 2. Factors influencing decision to request data through BioLINCC described by 195 survey respondents.

Note: Respondents were able to select multiple answers in response to this question.

Figure 3. Reasons why project using clinical research data from BioLINCC between 2007 and 2014 has not yet been completed (n=97).

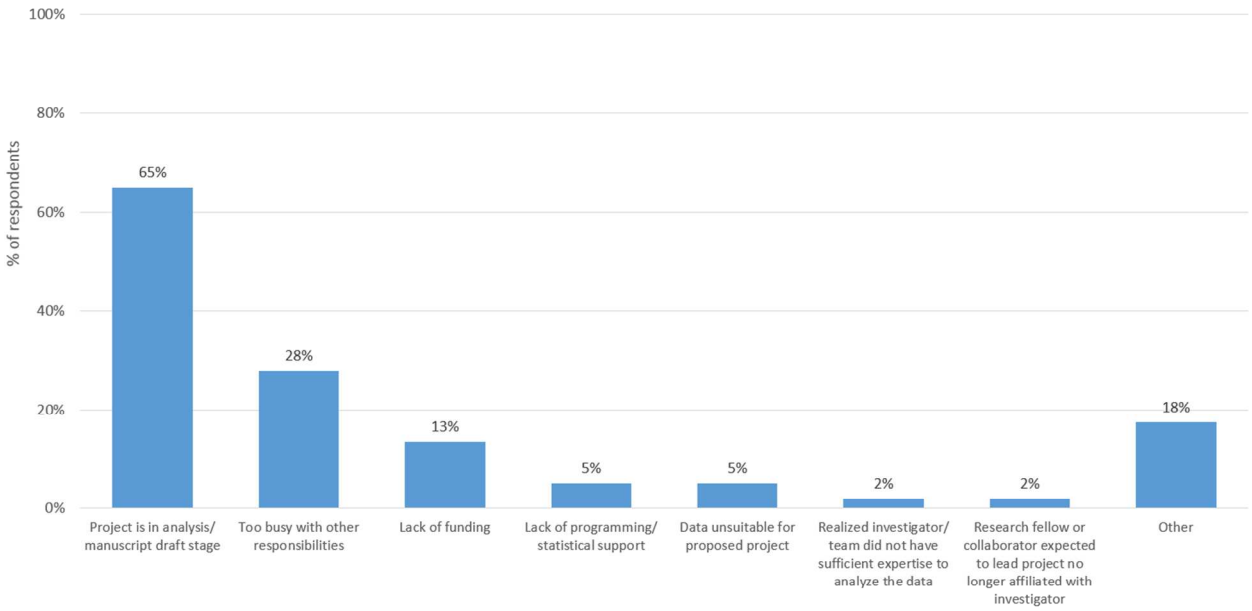

Fig 3. Reasons for incomplete project described by 97 survey respondents.

Note: Respondents were able to select multiple answers in response to this question.

Data Sharing through a NIH Central Database Repository:

A cross-sectional survey of BioLINCC users

APPENDIX

Table of Contents

BioLINCC Opt-in/Opt-out E-mail ..... page 2

Yale University Invitation E-mail ..... page 3

Survey ..... page 4

BioLINCC Opt-in/Opt-out E-mail

In an effort to better understand the benefits, experiences, issues, and barriers to investigators requesting data from data repositories, Yale University will be conducting a short survey of investigators that received data from the NHLBI Data Repository at any time between 2007 and 2014. The survey should take approximately 15 minutes to complete and will provide valuable information on the experiences of users of data repositories. Please note that your responses to the survey will be completely anonymous. NHLBI will not receive nor have any access to any of the survey responses. NHLBI may request specific tables to further the Institute's understanding of potential areas for improvements; however, your individual responses will remain anonymous.

At the conclusion of the survey and publication of findings, your contact information will be permanently removed from all Yale University systems.

Please respond with either a 'Yes' indicating that NHLBI has your permission to share your contact information only with Yale University and only for the purpose of carrying out the survey of NHLBI Data Repository investigators. No permission to share is implied for any other purpose with any other third party.

Or Respond with a 'No' indicating that you do not wish to participate in the survey.

Yale University Invitation E-mail

Subject: Yale Survey on Using Data from NHLBI's Data Repository (BioLINCC)

Yale University, in collaboration with the National Heart, Lung, and Blood Institute (NHLBI) of the NIH, is conducting a short survey of investigators that received data from the NHLBI Data Repository (BioLINCC) at any time between 2007 and 2014. This survey is intended to better understand the benefits, experiences, issues, and barriers to investigators requesting data from this repository.

After having the opportunity to use data from BioLINCC, we hope that you will consider providing feedback on this valuable resource so that efforts can be made to enhance its use by others.

Participation in the survey is voluntary and responses will be anonymized. This survey is expected to take no more than 15 minutes to complete and all participants will be automatically entered in a drawing to win one of five Amazon.com© gift certificates worth $100.

Please complete this survey by June 4, 2015.

Thank you for your participation!

Survey

Yale University is conducting a survey of investigators who have accessed clinical research data through the Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC), at the U.S. National Heart, Lung, and Blood Institute (NHLBI) of the NIH, from 2007 through 2014.

Specifically, we are interested in investigators' experiences with the data and perceptions of the value, importance, and challenges of data sharing initiatives such as this one.

Results from this project are intended to improve investigators' experience with BioLINCC, as well as to inform future clinical data sharing efforts, and are intended to be published in the biomedical literature.

There are no physical risks associated with this project. Participation is voluntary and responses will be anonymized. This survey is expected to take no more than 15 minutes to complete and all participants will be automatically entered in a drawing to win one of five Amazon.com© gift certificates worth $100. Completion of the survey indicates consent to participate.

Please contact us at jessica.ritchie@yale.edu or (203) 200-5346 if you have questions or concerns related to the survey.

I. Reasons for Data Request

1. For what primary research purpose(s) did you request data through NHLBI's Data Repository (BioLINCC)? (Please check all that apply)
   - To conduct an independent scientific study/studies
   - To conduct a pilot/preliminary analysis
   - To conduct an analysis with bio-specimens
   - To learn more about the BioLINCC data request process
   - Other (please specify): [Free text field]

2. Did any of the following influence your decision to request data through NHLBI's Data Repository (BioLINCC)? (Please check all that apply)
   - The study was closed and individual participant-level data were not available elsewhere
   - Original study investigators suggested BioLINCC as the appropriate data source
   - Difficulties in establishing a collaboration with original study investigators
   - Collecting data of similar size and scope was not feasible
   - Insufficient financial resources available for primary data collection
   - Insufficient time for primary data collection
   - Insufficient experience with primary data collection
   - Insufficient mentorship to support primary data collection
   - Unable to access similar data through home institution electronic medical records
   - Other (please specify): [Free text field]

3. How did you intend to use the data? (Please check all that apply) *[**BIFURCATION QUESTION**: If participants answer "To be combined", they will be presented with Question #4. If participants answer "As a standalone data source" or "Other", they will be presented with Question #6.]*
   - As a standalone data source

To be combined with other data sources from BioLINCC
To be combined with other data sources not from BioLINCC (i.e., non-BioLINCC studies, other public data)
Other (please specify): [Free text field]

4. When combining the requested data with other data sources, was the purpose of the project to conduct a meta-analysis? *[**NOTE:** Question #4 only asked of participants who answered "To be combined" to Question #3; **BIFURCATION QUESTION**: If participants answer "Yes", they will be presented with Question #5. If participants answer "No", they will be presented with Question #6.]*
   Yes
   No

5. For the meta-analysis, which of the following was planned? (Please check all that apply) *[**NOTE:** Question #5 only asked of participants who answered "Yes" to Question #4.]*
   Summary-level data meta-analysis
   Participant-level data meta-analysis
   Other (please specify): [Free text field]

6. What was your primary research objective for the data requested through NHLBI's Data Repository (BioLINCC)? (Please check all that apply)
   New research
   Replication research
   Other

7. You indicated that your primary research objective was New research. Which of the following further describe this objective? (Please check all that apply) *[**NOTE:** Question #7 only asked of participants who answered "New research" to Question #6.]*
   To examine secondary endpoints
   To examine subgroup populations
   To leverage the data for a research question unrelated to the original research design (i.e., examine lost-to-follow-up rates or endpoints used in clinical trials)
   To leverage the data to create a cohort for comparison to another study
   Other (please specify): [Free text field]

8. You indicated that your primary research objective was Replication research. Which of the following further describe this objective? (Please check all that apply) *[**NOTE:** Question #8 only asked of participants who answered "Replication research" to Question #6.]*
   Replicate the main study primary endpoint findings
   Replicate the main study secondary endpoint findings
   Replicate the main study subgroup findings (for primary and/or secondary endpoints)
   Other (please specify): [Free text field]

9. You indicated that your primary research objective was Other. Which of the following further describe this objective? (Please check all that apply) *[**NOTE:** Question #9 only asked of participants who answered "Other" to Question #6.]*
   - Statistical methods research
   - Epidemiological research
   - Preliminary research to be used as part of a grant proposal
   - Other (please specify): [Free text field]

10. Did your research focus on a medical product intervention (i.e., drug, biologic, medical device? *[**BIFURCATION QUESTION**: If participants answer "Yes", they will be presented with Question #11. If participants answer "No", they will be presented with Question #12.]*
   - Yes
   - No

11. What was the focus of your primary research question? (Please check all that apply) *[**NOTE:** Question #11 only asked of participants who answered "Yes" to Question #10.]*
   - Efficacy
   - Safety
   - Pharmacodynamics
   - Other (please specify): [Free text field]

II. Interactions with Original Study Investigators

12. Prior to or after requesting data through NHLBI's Data Repository (BioLINCC), did you contact the original study investigators to obtain the data? *[**BIFURCATION QUESTION**: If participants answer "Yes", they will be presented with Question #13. If participants answer "No", they will be presented with Question #16.]*
   - Yes
   - No

13. Did the original study investigators approve your data request? [**NOTE:** *Question #13 only asked of participants who answered "Yes" to Question #12;* **BIFURCATION QUESTION**: *If participants answer "Yes", they will be presented with Question #14. If participants answer "No", they will be presented with Question #15.]*
   - Yes
   - No

14. If the original study investigators approved your data request, for what reason(s) did you also request data through BioLINCC? (Please check all that apply) *[**NOTE:** Question #14 only asked of participants who answered "Yes" to Question #13.]*
   - Wanted to validate data made available by original study investigators
   - Original study investigators required co-authorship to make data available
   - Original study investigators required control of publication to make data available

Original study investigators required control of study design to make data available
Original study investigators required control of data analysis to make data available
Data made available by original study investigators had no or poor accompanying documentation
Data made available by original study investigators were poorly organized and could not be prepared for analysis
More straightforward to access data through BioLINCC
Other (please specify): [Free text field]

15. What reasons did the original study investigators provide for not approving your data request? (Please check all that apply) *[**NOTE:** Question #15 only asked of participants who answered "No" to Question #13.]*
    No interest in collaborating with external investigators
    Data cannot be made available to external investigators because original human subject consent forms do not allow
    Data cannot be made available to external investigators because of intellectual property issues
    Data cannot be made available to external investigators because of patient confidentiality issues
    BioLINCC was suggested as the appropriate data source
    A reason was not provided
    Other (please specify): [Free text field]

16. Prior to or after requesting data through BioLINCC, did you contact the original study investigators to request collaboration? *[**BIFURCATION QUESTION**: If participants answer "Yes", they will be presented with Question #17. If participants answer "No", they will be presented with Question #20.]*
    Yes
    No

17. For what reason(s) did you request collaboration with the original study investigators? (Please check all that apply) *[**NOTE:** Question #17 only asked of participants who answered "Yes" to Question #16.]*
    Needed additional data that were not included in the files provided by BioLINCC
    Needed additional statistical expertise due to data complexity
    Needed additional content expertise due to study design complexity
    Needed additional clinical expertise related to study question
    Wanted to work with original study investigators
    Other (please specify): [Free text field]

18. Did the original study investigators accept your request for collaboration? *[**NOTE:** Question #18 only asked of participants who answered "Yes" to Question #16.]*
    Yes
    No

19. What reasons did the original study investigators provide for not accepting your request to collaborate? (Please check all that apply) [***NOTE:*** *Question #19 only asked of participants who answered "No" to Question #18.]*
    - A reason was not provided
    - Original study investigators were too busy
    - Original study investigators felt the research question was low priority
    - Original study investigators did not have funds to support collaboration
    - Other (please specify): [Free text field]

III. Data Repository Experience

20. How did you learn about the NHLBI's Data Repository (BioLINCC)? (Please check all that apply)
    - Internet search
    - Communications with NHLBI
    - Colleagues/other investigators
    - Directed to BioLINCC by study investigators
    - Other (please specify): [Free text field]

21. Is there anything you would have liked to have known prior to accessing the BioLINCC data?
    - No
    - Yes (please provide further details): [Free text field]

22. Were the data you received from BioLINCC suitable for your originally proposed project? [***BIFURCATION QUESTION****: If participants answer "Yes", they will be presented with Question #24. If participants answer "No", they will be presented with Question #23.*]
    - Yes
    - No

23. For what reason(s) was the data unsuitable for your proposed project? (Please check all that apply) [***NOTE:*** *Question #23 only asked of participants who answered "No" to Question #21.*]
    - Data had no or poor accompanying documentation; could not determine if data were suitable
    - Data were too complicated to use; could not determine if data were suitable
    - Data were poorly organized; could not be adequately prepared for analysis
    - Data had too many missing values; could not be adequately prepared for analysis
    - Proposed main outcome variable was not available in data
    - Main outcome variable was not available in data at the time points proposed for study
    - Proposed main independent variable was not available in data
    - Other (please specify): [Free text field]

24. Please consider your experience using the data you received from BioLINCC. Do you agree or disagree with the following statement: The documentation and data dictionaries (i.e., meta-data) received from BioLINCC were useful. [***BIFURCATION QUESTION:*** *If participants answer*

"Somewhat disagree" or "Strongly disagree", they will be presented with Question #25. Otherwise, they will be presented with Question #26.]

Strongly agree
Somewhat agree
Somewhat disagree
Strongly disagree

25. You chose "Somewhat disagree" or "Strongly disagree." Please briefly explain your answer. [***NOTE:*** *Question #25 only asked of participants who answered "Somewhat disagree" or "Strongly disagree" to Question #24.]*
    [Free text field]

III. Project Details

26. Has your project been completed? *[**BIFURCATION QUESTION:*** If participants answer "No", they will be presented with Question #27. If participants answer "Yes", they will be presented with Question #31.]
    Yes
    No

27. For what reason(s) was the project not completed? (Please check all that apply) *[**NOTE:** Question #27 only asked of participants who answered "No" to Question #26.]*
    Project is in analysis/manuscript draft stage
    Data unsuitable for proposed project
    Realized investigator/team did not have sufficient expertise to analyze the data
    Lack of funding
    Lack of programming/statistical support
    Too busy with other responsibilities
    Research fellow or collaborator expected to lead project no longer affiliated with investigator
    Investigator no longer active in clinical research
    Others published same/similar work on same/similar data
    Other (please specify): [Free text field]

28. Do you plan to complete the project? *[**NOTE:** Question #28 only asked of participants who answered "No" to Question #26. **BIFURCATION QUESTION:*** If participants answer "No", they will be presented with Question #29. If participants answer "Yes", they will be presented with Question #31.]
    Yes
    No

29. Was there an issue with the BioLINCC data that prevented you from completing the project? *[**NOTE:** Question #29 only asked of participants who answered "No" to Question #28. **BIFURCATION QUESTION:*** If participants answer "Yes", they will be presented with Question #30. If participants answer "No", they will be presented with Question #39.]
    Yes

No

30. Please briefly explain your answer. *[**NOTE:** Question #30 only asked of participants who answered "Yes" to Question #29.]*
    [Free text field]

31. Does your completed or anticipated final project differ from your pre-specified project? *[**NOTE:** Question #31 only asked of participants who answered "Yes" to either Question #26 or #28*. ***BIFURCATION QUESTION:*** If participants answer "Yes", they will be presented with Question #32. If participants answer "No", they will be presented with Question #33. **NOTE:** If participants answer "No" to #26 and "No" to #31, they will be presented with #38]
    Yes
    No

32. In what ways does your completed or anticipated final project differ from your pre-specified project? (Please check all that apply) *[**NOTE:** Question #32 only asked of participants who answered "Yes" to Question #31.]*
    Modified planned data source by combining data received from BioLINCC with other data sources
    Modified planned data source by not combining data received from BioLINCC with other data sources
    Modified study sample
    Modified primary endpoints
    Modified secondary endpoints
    Modified selection of main independent variables
    Modified statistical analysis plan
    Other (please specify): [Free text field]

33. Was your research project published? [***NOTE:** Question #33 only asked of participants who answered "Yes" to Question #26*. ***BIFURCATION QUESTION:*** If participants answer "Yes", they will be presented with Question #34. If participants answer "No", they will be presented with Question #36.]
    Yes
    No

34. In what format was your research project published? (Please check all that apply) [***NOTE:** Question #34 only asked of participants who answered "Yes" to Question #33*.]
    Original research article in a peer-reviewed biomedical journal
    Systematic review/meta-analysis in a peer-reviewed biomedical journal
    Non-systematic review article in a peer-reviewed biomedical journal
    Commentary / viewpoint / editorial in a peer-reviewed biomedical journal
    Letter in correspondence in a peer-reviewed biomedical journal
    Weblog post or other on-line forum
    Self-published
    Other (please specify): [Free text field]

35. Please provide the publication citation and PubMed ID (if applicable) or other citation (such as web address). [***NOTE***: *Question #35 only asked of participants who answered "Yes" to Question #33.*]

    [Free text field]

36. When attempting to publish your research, were any of the following concerns raised by editors or peer reviewers during the peer-review process? (Please check all that apply)

    I did not attempt to publish
    No substantive concerns were raised beyond minor comments and suggestions to clarify/improve the research
    Concern that I was not one of the original study investigators
    Concern about the original study design that I could not address
    Concern about my research project design that I could not address without additional data
    Concern about my research project design unrelated to the data
    Concern about my research methodology and analysis
    Concern about the importance of my research
    Other (please specify): [Free text field]

37. How many research projects did you complete, or do you plan to complete, through your single request? [***NOTE:*** *Question #38 only asked of participants who answered "Yes" to either Question #26 or #28.*]

    One
    Two
    Three
    Four or more

38. Did using data from BioLINCC aid in any future grant applications? (Please check all that apply)

    Yes, use of the BioLINCC data established a publication record that was then included in a grant application
    Yes, use of the BioLINCC data furthered the understanding of questions of particular interest/identified gaps that then served as the basis of a grant application
    Yes, other (please specify): [Free text field]
    No

39. What was the primary funding source used to support this project?

    Self-funded
    NIH
    Non-NIH Federal
    Non-Profit Organization or Foundation in US
    Industry
    Non-US Government or Organization
    Other (please specify): [Free text field]

40. What additional funding source(s) were used to support this project? (Please check all that apply)

    Self-funded

NIH
Non-NIH Federal
Non-Profit Organization or Foundation in US
Industry
Non-US Government or Organization
Other (please specify): [Free text field]
None

IV. Requestor Demographics

41. Which of the following best classifies your primary employer at the time when you requested data through BioLINCC?
    Academic Institution
    Private Industry
    Non-Profit Organization
    For-Profit Hospital
    Government
    Other (please specify): [Free text field]

42. How would you classify your career status with respect to clinical or epidemiological research at the time when you requested data through BioLINCC?
    In training (< 3 years of active engagement in clinical research, still receiving formative training in research methods)
    Early stage career (3-10 years of active engagement in clinical research)
    Established in the field (> 10 years of active engagement in clinical research)

43. Have you ever been closely involved (as Principal or Co-Investigator) in the conduct of a randomized controlled trial?
    Yes
    No

44. Have you ever deposited clinical trial data in the BioLINCC repository?
    Yes
    No

45. Please indicate your age range.
    34 years or younger
    35-49 years
    50-64 years
    65 years or older
    Prefer not to answer

46. Please indicate your gender.
    Male
    Female
    Prefer not to answer

47. Please indicate your ethnicity.

Hispanic or Latino
Not Hispanic or Latino
Prefer not to answer

48. Please indicate your race.
    American Indian or Alaska Native
    Asian
    Black or African American
    Native Hawaiian or Other Pacific Islander
    White
    Other (please specify): [Free text field]
    Prefer not to answer

49. Thank you for completing this survey. To be eligible for entry into the Amazon.com gift certificate drawing, please provide your email address. Your email address will be kept separate from your survey responses in order to ensure anonymity.

50. Would you like to receive a notification when the results of this study are published? If so, please re-enter your email address. Your email address will be kept separate from all other responses to ensure confidentiality.

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	3-4
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	5-6
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	6-7
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6-7
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	6-7
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	8-9
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	8-9
Bias	9	Describe any efforts to address potential sources of bias	9-10
Study size	10	Explain how the study size was arrived at	9-10
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	8-10
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	9-10
		(*b*) Describe any methods used to examine subgroups and interactions	n/a
		(*c*) Explain how missing data were addressed	n/a
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	n/a
		(*e*) Describe any sensitivity analyses	n/a
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	10
		(b) Give reasons for non-participation at each stage	10
		(c) Consider use of a flow diagram	Figure 1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	10, 25-28
		(b) Indicate number of participants with missing data for each variable of interest	10, 27-28
Outcome data	15*	Report numbers of outcome events or summary measures	11-13
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	11-13
		(*b*) Report category boundaries when continuous variables were categorized	n/a
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	n/a
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	n/a
Discussion			
Key results	18	Summarise key results with reference to study objectives	13-15
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	16-17
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	13-17
Generalisability	21	Discuss the generalisability (external validity) of the study results	13-15
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	19

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Data Sharing through a NIH Central Database Repository: A cross-sectional survey of BioLINCC users

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012769.R1
Article Type:	Research
Date Submitted by the Author:	29-Jul-2016
Complete List of Authors:	Ross, Joseph; Yale University School of Medicine, Internal Medicine Ritchie, Jessica; Yale-New Haven Hospital, Center for Outcomes Research and Evaluation Finn, Emily; Yale-New Haven Hospital, Center for Outcomes Research and Evaluation Desai, Nihar; Yale-New Haven Hospital, Center for Outcomes Research and Evaluation Lehman, Richard; Oxford University, Primary Health Care Krumholz, Harlan; Yale University, Medicine Gross, Cary; Yale School of Medicine, Section of General Internal Medicine
Primary Subject Heading:	Medical publishing and peer review
Secondary Subject Heading:	Health policy, Health services research, Evidence based practice, Ethics
Keywords:	Clinical trials < THERAPEUTICS, STATISTICS & RESEARCH METHODS, PUBLIC HEALTH

Data Sharing through a NIH Central Database Repository:

A cross-sectional survey of BioLINCC users

Joseph S Ross, MD, MHS, *associate professor of medicine*[1,2,3,4]; Jessica D Ritchie, MPH, PMP, *research coordinator*[1]; Emily Finn, MPH, *research associate*[2]; Nihar R Desai, MD, *assistant professor of medicine*[1,5]; Richard L Lehman, MD, *senior advisory fellow*[1,6]; Harlan M Krumholz, MD, SM, *Harold H Hines Jr, professor of medicine and epidemiology and public health*[1,3,4,5]; Cary P Gross, MD, *professor of medicine*[2,3,7].

[1] Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, CT 06510, USA;

[2] Section of General Internal Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT 06510, USA;

[3] Robert Wood Johnson Foundation Clinical Scholars Program, Department of Internal Medicine, Yale School of Medicine, New Haven, CT 06510, USA;

[4] Department of Health Policy and Management, Yale School of Public Health, New Haven, CT 06520, USA;

[5] Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT 06510, USA;

[6] UK Cochrane Center, Oxford, England;

[7] Cancer Outcomes, Public Policy, and Effectiveness Research Center, Yale Cancer Center, New Haven, CT 06510, USA.

Note: Ms. Finn was affiliated with the Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, during the time the work was conducted.

Correspondence: Joseph S. Ross, MD, MHS, Section of General Internal Medicine, Yale University School of Medicine, P.O. Box 208093, New Haven, CT 0520, Phone: (203) 785-2987, Fax: (203) 737-3306, email: joseph.ross@yale.edu.

Word count (text only): 4294 **Tables:** 2 **References:** 25

Abstract word count: 324 **Figures:** 3

Key words: Clinical trials; Data sharing; Open science.

ABSTRACT

Objective

To characterize experiences with using clinical research data shared through the National Institute of Health's Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC) clinical research data repository, along with data recipients' perceptions of the value, importance, and challenges with using BioLINCC data.

Design and Setting

Cross-sectional web-based survey.

Participants

All investigators who requested and received access to clinical research data from BioLINCC between 2007 and 2014.

Main Outcome Measures

Reasons for BioLINCC data request, research project plans, interactions with original study investigators, BioLINCC experience, and other project details.

Results

There were 536 investigators who requested and received access to clinical research data from BioLINCC between 2007 and 2014. Of 441 potential respondents, 195 completed the survey (response rate=44%); 89% (n=174) requested data for an independent study, 17% (n=33) for pilot/preliminary analysis. Commonly cited reasons for requesting data through BioLINCC were feasibility of collecting data of similar size and scope (n=122) and insufficient financial resources for primary data collection (n=76). For 95% of respondents (n=186), a primary research objective was to complete new research, as opposed to replicate prior analyses. Prior to requesting data from BioLINCC, 18% (n=36) of respondents had contacted the original study investigators to obtain data, whereas 24% (n=47) had done so to request collaboration. Nearly all (n=176; 90%) respondents found the data to be suitable for their proposed

project; among those who found the data unsuitable (n=19; 10%), cited reasons were data too complicated to use (n=5) and data poorly organized (n=5). Half (n=98) of respondents had completed their proposed projects, of which 67% (n=66) have been published.

Conclusions

Investigators were primarily using clinical research data from BioLINCC for independent research, making use of data that would otherwise have not been feasible to collect.

Article Summary: Strengths and Limitations of this Study

- Data sharing policies are increasingly promoted and being adopted by research funders to improve access to clinical trial data to inform evidence-based practice. The National Institute of Health's Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC) has been actively sharing data from its clinical research data repository for more than 10 years.
- In the first survey of the experiences of investigators who have requested and been approved to use data from BioLINCC, we found that users were primarily focused on conducting independent research studies, making use of data that would otherwise have not been feasible to collect, because of both insufficient time and resources.
- We also found that shared data from BioLINCC could be used to successfully pursue clinical research; 90% of BioLINCC users found the data to be suitable, half had completed their research projects thus far, and two-thirds had published their findings.
- Our study of user experiences with BioLINCC offer important insights for newly initiated and on-going clinical trial data sharing efforts and illustrate the potential and value of data sharing for the broader scientific field, as well as the challenges that remain to be overcome.
- Our study is limited by a low response rate and may have been affected by recall bias and social desirability bias, perhaps suggesting that our findings overestimate the perceived value of BioLINCC data and its usability for the broader scientific community.

Over the past 5 years, several major research funders, including the U.S. National Institutes of Health (NIH), the U.S. Patient-Centered Outcomes Research Institute, the U.K. Medical Research Council, and the Bill and Melinda Gates Foundation, as well as private industry,[1] have adopted policies supporting or mandating clinical research data sharing. In January 2015, the Institute of Medicine of the U.S. National Academies further supported these efforts with its report, "Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risks," recommending that stakeholders foster a culture in which data sharing is the expected norm and commit to responsible strategies aimed at maximizing benefits, minimizing risks, and overcoming challenges of sharing clinical trial data.[2] In January 2016, the International Committee of Medical Journal Editors issued a proposal to require authors to share with others the deidentified individual-patient data underlying the results presented in the article no later than 6 months after publication as a condition of consideration for publication of a clinical trial report in our member journals.[3]

In response to these new policies and proposals, funded investigators will increasingly be asked to prepare and make collected data available to other investigators with whom they are not collaborating so that the second can pursue independent research. To support these efforts and inform developing policies, a number of prior studies have examined the willingness of clinical trial investigators to share clinical research data, generally finding broad support, and characterized anticipated challenges to and concerns with data sharing.[4-11] However, few studies have focused on the investigators who have actually received deidentified individual-patient data from a centralized data sharing platform, in order to understand their perspectives regarding challenges encountered with requesting and using the data, and disseminating findings.

While most of these data sharing efforts have been relatively newly established, the U.S. National Heart, Lung, and Blood Institute (NHLBI) of the NIH established a formal data repository in 2000, now managed by the Biologic Specimen and Data Repository Information Coordinating Center

(BioLINCC), to facilitate access to, maximize the scientific value of, and promote the availability and use of the biorepository, data repository and other NHLBI-funded population-based biospecimen and data resources by investigators worldwide.[12 13] The BioLINCC data repository includes individual level data on more than 580,000 participants from over 110 Institute supported clinical trials and observational studies, beginning as far back as the 1980s. Each data set is prepared independently by the NHLBI-funded investigator to comply with specific requirements and data standards, with oversight by BioLINCC, including provision of baseline, interim visit, ancillary study and outcome data for clinical trials and provision of all examination and ancillary study data, along with follow-up information, for epidemiology studies. As BioLINCC has been actively sharing data for more than a decade and currently receives over 100 requests for clinical trial and other prospective cohort clinical data per year (ref: personal communication, Sean Coady, NHLBI data repository manager), there is an opportunity to learn from data users' experiences to inform clinical data sharing efforts. Accordingly, we surveyed all investigators who requested and received access to clinical research data from BioLINCC between 2007 and 2014. We specifically sought to understand their experiences with clinical research data sharing and status of their research project, as well as perceptions of the value, importance, and challenges of accessing data through BioLINCC.

METHODS

Study Sample and Design

We conducted a cross-sectional survey from May to August 2015 of all investigators who requested and received access to clinical research data from BioLINCC between 2007 and 2014. This time period was chosen to ensure a contemporaneous sample of investigators whose contact information was less likely to have changed over ensuing years. In accordance with NIH policy, BioLINCC provided our study team with a list of investigators who had requested and received access using a

public e-mail address; contact information was available for the lead investigator who was responsible for the BioLINCC request, not each member of the study team. For investigators who had requested and received access using a *private e-mail address*, BioLINCC first sent an opt-in/opt-out e-mail in May 2015 asking if they would be willing to participate in the survey (**Appendix**). Non-respondents were sent two follow-up requests by e-mail; those that did not respond by the end of the third week were considered to have opted-out. BioLINCC subsequently provided our study team with a list of those investigators who opted-in.

In addition to contact information, BioLINCC provided our study team with information on the following for all investigators who had requested and received access to clinical research data: lead investigator location, affiliation with an academic institution or for-profit organization, and total number of requests ever submitted to BioLINCC, as well as the request year, the number of data sets requested, and self-reported availability of external funding to support the research project using the requested data.

In May 2015, the Yale team sent all potential survey respondents an initial e-mail to describe the purpose of the study, request their participation, and provide a link to the survey; three follow-up requests were sent by e-mail over the course of June 2015. Non-respondents were contacted by telephone to solicit their participation up to twice per week, but no more than once per day, until one contact was made. In July 2015, Internet searches to update contact information for non-respondents were conducted. For all non-respondents for whom updated contact information was identified, the initial survey email was sent, followed by three follow-up requests.

Invitations to participate did not reference a specific hypothesis of the study, but stated that investigator participation would further the understanding of investigators' experience with BioLINCC and inform future clinical trial data sharing efforts (**Appendix**). Participation was voluntary and included an opportunity to win one of five $100 gift certificates for Amazon. All internet-based responses were

collected using a Web-based survey platform (Qualtrics Labs, Provo, UT). Approval from the Yale University School of Medicine Human Research Protection Program was obtained prior to study conduct and consent was considered to be implied when participants completed the online survey.

Survey Instrument Development

The design of our 50-item survey instrument was informed by previously published surveys,[4 5] a review of the literature on clinical trial data sharing, and discussion with multiple experts and stakeholders, including representatives from NHLBI and academic investigators. Experts recommended survey topics that they considered to be compelling for the field of data sharing and re-use of data. The survey was pre-tested with six medical students and staff at the Yale-New Haven Hospital Center for Outcomes Research and Evaluation (New Haven, CT) and modified iteratively to improve clarity, face validity, and content validity. Adaptive questioning was used to decrease response burden. Items were presented in multiple response, Likert scale, and open-ended formats; many of the multiple response questions enabled respondents to select multiple answers. The complete instrument is provided within the **Appendix**.

Survey Domains

Reasons for Data Request and Planned Research Project

We used multiple response and yes/no questions to assess investigators' primary research purpose and reasons for requesting data from BioLINCC. Multiple response questions were also used to determine the primary research objective, funding used to support the project, and other details of the planned research project. Knowing what these clinical research data are being used for will help tailor future data sharing efforts to the needs of investigators.

Interactions with Original Study Investigators

We used yes/no questions to determine whether original study investigators were contacted prior to or after requesting data through BioLINCC to obtain the data or to collaborate. These were followed by multiple response questions to determine why collaborations were sought, whether the requests for data or collaboration were approved, and reasons for not approving. Answers to these questions could potentially demonstrate the value of a data resource such as BioLINCC.

BioLINCC Experience

Multiple response, yes/no, and Likert-type questions were used to obtain information regarding investigator's experience using BioLINCC, including whether the data were suitable and useful for their project. Knowledge gained from these questions can help to improve BioLINCC and other data sharing efforts.

Project Details

We used multiple response and yes/no questions to characterize the completion stage of investigators' projects. For those that did not complete their project, multiple response and yes/no questions were used to ascertain reasons why the project was incomplete. For those with completed projects, we used multiple response and yes/no questions to determine whether the final project differed from the pre-specified project as well as to obtain publication information. Multiple choice and multiple response questions were used to identify any funding sources and whether using the data from BioLINCC aided in any future grant applications. It is important to demonstrate not only that these data are being requested, but that they are also being used to potentially generate new knowledge to advance science and public health.

Requestor Demographics

Respondents were asked to characterize their primary employer and career status using multiple choice questions, including whether they had ever been closely involved (as Principal or Co-Investigator) in the conduct of a randomized controlled trial and/or ever deposited clinical trial data in the BioLINCC repository. Respondent sociodemographic characteristics, including age, gender, and ethnicity, were also collected. While these characteristics were collected for descriptive purposes only, age, along with the professional characteristics collected, are of importance to demonstrate the value of the availability of BioLINCC data to investigators who are in certain stages of their career.

Patient Involvement

Patients were not involved in the design or conduct of this study. Results will be directly disseminated via email to all individuals invited to participate in the survey upon publication.

Statistical Analysis

To compare characteristics of survey respondents and non-respondents, we used two-sided Chi-square tests and Fisher Exact tests when appropriate with a type 1 error level of 0.05. Next, we conducted descriptive analyses of the reasons for requesting data from BioLINCC, prior interactions with original trial investigators, experience using BioLINCC, and project details, as well as respondent demographic characteristics. Data were analyzed using JMP Pro Version 11.2.0 (SAS Institute Inc., Cary, NC).

RESULTS

There were 536 investigators who requested and received access to clinical research data from BioLINCC between 2007 and 2014 (**Figure 1**). Investigators for which a public e-mail address was not

available were sent an opt-in/opt-out letter (n=74); 23 opted in, 3 opted out, 7 could not be reached, and 41 were not responsive. Survey participation requests were thus sent to 485 eligible respondents, 44 of whom were subsequently excluded due to the following reasons: invalid contact information (n=31), the investigator had no recollection of requesting the data (n=5), or the data had been requested by someone other than the investigator (n=8). Of the remaining 441 respondents, 195 completed the survey, yielding a survey response rate of 44.2%. However, of the 536 total investigators who requested and received access to clinical research data from BioLINCC, 195 completed the survey (response rate of 36.3%).

Survey respondents did not differ from non-respondents with respect to investigator location, affiliation with an academic institution or for-profit organization, and total number of requests ever submitted to BioLINCC, as well as the number of data sets requested (P values ≥ 0.10; **Table 1**). However, respondents were more likely than non-respondents to have requested data more recently (P=0.004) and to have self-reported external funding to support the research project (P=0.009).

Half of survey respondents were between the ages of 35 and 49 years old (n=97; 50%), while 59% were male (n=116), 68% were white (n=133), and 90% identified as not Hispanic/Latino (n=175; **Table 2**). The vast majority of respondents were primarily employed by an academic institution (n=165; 85%) and 78% (n=152) have been engaged in clinical research for at least three years. While 42% (n=82) had been closely involved in the conduct of a randomized controlled trial, only 3% (n=5) had ever deposited data in the BioLINCC repository.

Reasons for Data Request

Overall, respondents' motivations for requesting data from BioLINCC were largely focused on using the data to conduct and disseminate new research studies, as 89% (n=174) indicated that data were requested for an independent study, 17% (n=33) to use the data for pilot/preliminary analysis. For

63% (n=122) of respondents, the decision to request data through BioLINCC was influenced by the belief that collecting data of similar size and scope was not feasible, while insufficient financial resources for primary data collection (n=76; 39%), individual participant-level data being unavailable elsewhere (n=71; 36%), and insufficient time for primary data collection (n=64; 33%) were also commonly cited reasons for requesting data through BioLINCC (**Figure 2**).

Planned Research Project

Respondents largely (n=149; 76%) planned research projects that used the requested BioLINCC data as a standalone data source for at least one project, while 43% (n=83) planned to combine the data with other data sources; of these, 27% (n=22) planned to conduct a meta-analysis. Nearly all respondents (n=186; 95%) indicated that at least one of their primary research objectives was to complete new research, whereas only 7 (4%) had a primary research objective solely to replicate prior analyses. Of those pursuing new research, 56% (n=104) planned to leverage the data for a research question unrelated to the original research design, while 40% (n=74) planned to examine subgroup populations and 32% (n=60) planned to examine secondary endpoints.

Only 13% (n=26) of respondents indicated that the focus of their research was a medical product or intervention; of these, 73% (n=19) planned analyses to examine product / intervention efficacy, 54% (n=14) safety. Finally, 52% (n=102) of respondents had funding to support the research project, most commonly from the NIH (n=44; 23%), whereas 43% (n=84) primarily self-funded the research project.

Interactions with Original Study Investigators

Fewer than one in five (n=36; 18%) respondents indicated that they had contacted the original study investigators to obtain data prior to requesting the data from BioLINCC; among these, 44% (n=16) reported that the original study investigator approved their request and these investigators most

commonly requested access to the data from BioLINCC anyway because the process to access data was more straightforward through BioLINCC (n=11). Among the 20 (56%) respondents who indicated that the original study investigator denied their request, the most common response given by the original investigator was to direct the respondent to BioLINCC (n=11; 55%).

Nearly one-quarter of respondents (n=47; 24%) indicated that they contacted the original study investigator to request collaboration, most commonly because of an interest in working with the original study investigators (n=23) and need for additional content expertise due to study design complexity (n=20). Of the respondents who requested collaboration, two-thirds (n=31; 66%) indicated that the request was accepted.

Data Repository Experience

Nearly all respondents indicated satisfaction with the data available through BioLINCC and that they were suitable for their originally proposed project (n=176; 90%). Among the 19 (10%) respondents who indicated that the data were not suitable, the two most commonly cited reasons were that the data were too complicated to use, preventing them from determining whether the data were suitable (n=5); and that the data were poorly organized, preventing adequate preparation for analysis (n=5).

Research Project Details

Half of all respondents (n=98; 50%) reported that their projects have been completed, of which 67% (n=66) have been published. Respondents who had requested data prior to 2012 were more likely to have completed their project when compared with those who had requested data in 2012 or afterwards (73% versus 44%; p=0.008). However, among those who completed their project, rates of publication did not differ among those who had requested data prior to 2012 and those who had requested data in 2012 or afterwards (63% versus 69%; p=0.57). Of those who have completed their

research, 48% (n=47) indicated that no substantive concerns were raised about the use of data from BioLINCC during the peer-review process, while 8% indicated that concerns were raised about research methodology and analysis (n=8), 7% about the original study design that the investigator could not address (n=7), and 6% about their research project design that they could not address without additional data (n=6).

Of the 97 respondents (50% of total) who have not yet completed their proposed projects, 84% (n=81) explained that they planned to complete their project; 65% (n=63) indicated that their project is in analysis/manuscript draft phase, while 28% (n=27) explained that they have thus far been too busy with other responsibilities to complete the research project using the data from BioLINCC and 13% (n=13) reported that lack of funding to support the project was a problem (**Figure 3**). Sixteen investigators explained that they did not intend to complete their project, most often because the age of the data made the project now less relevant or because of data issues, such as missing values for the variable of interest.

Of the 179 respondents who already completed or planned to complete their proposed project, 54% (n=96) reported that there would be one research project resulting from their single request for data from BioLINCC, 23% (n=42) reported two, and 23% (n=41) reported three or more. In addition, 15% (n=27) of respondents who have completed or planned to complete their project indicated that their completed/anticipated final project differed from their pre-specified project; the most commonly modified aspects were the statistical analysis plan (n=18) and the selection of the main independent variables (n=12).

DISCUSSION

In this survey of investigators who had requested and received access to clinical research data from BioLINCC between 2007 and 2014, the vast majority had requested the data in order to conduct

independent research projects, primarily because collecting data of similar size and scope was not feasible, due to both insufficient time and resources. Half of the investigators had completed their research projects thus far, two-thirds of which published their findings, and among those investigators whose projects were incomplete, two-thirds were actively engaged in analysis or manuscript preparation. These findings offer important insights for newly initiated and on-going clinical trial data sharing efforts and illustrate the potential and value of data sharing for the broader scientific field, as well as the challenges that remain to be overcome.

First, the BioLINCC experience suggests that when clinical research data are made available to investigators, there is likely to be interest in using the data for independent research projects. There are currently 654 publications associated with the data repository available through BioLINCC.[14] This large number of publications suggests that these data are being used by investigators, better maximizing the NHLBI investment in and scientific value of clinical research data. Many investigators responding to our survey noted that collecting data of similar size and scope was not feasible, or that they had insufficient financial resources or time for primary data collection, justifying the need to request data from BioLINCC for their research.

Second, the BioLINCC experience suggests that clinical data can be collected by one set of investigators and made available to another set of investigators who, for the most part, can use it to successfully pursue an independent research project. While some surveyed investigators noted challenges in using the data made available through BioLINCC, 90% found the data to be suitable for their originally proposed project, even without input from the original research team. Few reported that the data were too complicated to use, preventing them from determining whether the data were suitable, or that the data were poorly organized.

Finally, the research enterprise is not optimally efficient, and the BioLINCC experience reflects this short-coming. In aggregate, more than 100 research projects were completed as a result of

respondent investigators using data made available through BioLINCC. However, despite all investigators having received data from BioLINCC at no cost, only half of investigators who had received data had completed their research projects thus far. While many more continue to work on their projects and intend to complete their work, the investment by NHLBI to make these data available should be matched by the effort of investigators to ensure that the projects are completed. Moreover, even among completed projects, only two-thirds were published. While BioLINCC maintains an updated list of publications that have resulted from use of this shared data,[14] mechanisms should be established to ensure that results from research made possible through data sharing are publicly disseminated, either through publication or through a results reporting initiative similar to ClinicalTrials.gov.

For the potential and value of data sharing to be fully realized, more needs to be accomplished. Part of the success of BioLINCC may be attributed to the NHLBI policy that supported studies with direct costs equal to or greater than $500K in any 1 year and identified as being of high programmatic interest, along with co-operative agreements with 500 or more participants are required to submit data as part of the grant award.[13] This policy establishes clear expectations for data sharing, so that data can be properly organized and de-identified and supportive documentation and materials prepared in anticipation of submitting data to BioLINCC. However, it's not clear whether this policy allows researchers to budget resources for this work. Currently, the NIH is seeking ways to broaden data sharing efforts across its institutes;[15] to enhance the likelihood of success of data sharing efforts, it should be clarified whether NIH-granted independent research funds can be used to prepare collected data for sharing through initiatives such as BioLINCC.

Similarly, financial support for investigators to use clinical research data that are being shared and made available would enhance efforts. Forty percent of investigators using data from BioLINCC had self-funded their research efforts, while an eighth were relying on funding from the NIH. However, among surveyed investigators who had not yet completed their proposed projects, lack of funding to

support the project was a commonly cited problem. Without financial support, efforts to share data are likely to fail to achieve their potential,[2] even despite the strong policies and proposals in favor of data sharing from other research funders, the Institute of Medicine, and the International Committee of Medical Journal Editors.

There are important limitations of our study to consider. First, only 44% of potentially eligible respondents completed our survey, perhaps suggesting that our findings overestimate the perceived value of BioLINCC data and its usability for the broader scientific community. Individuals who chose not to respond to our survey may have found the data to be more problematic and less useful than those who responded. Furthermore, even among respondents, our findings may have been biased by recall bias, including an inability to remember using the data made available by BioLINCC, and social desirability,[16 17] as respondents may have been less likely to self-report experiences and project completion plans that may be negatively perceived by others. In addition, there were a few observed differences between survey respondents and non-respondents. Because we would expect that investigators who made more recent requests and who had secured external funding to support the research project would be more likely to remain enthusiastic about the project and to complete it, our findings may be biased toward higher project completion rates. However, our response rate compares favorably with other surveys of physicians and investigators,[4 18-20] perhaps reflecting that we used several mechanisms to prospectively improve response rates, including a web-based survey platform for ease of completion, we employed several reminder contacts, including three e-mails and at least one telephone contact and we offered financial incentives for participation .

Second, our study was limited to investigators who had received data from BioLINCC and our findings may not be applicable to the experience of investigators obtaining data from other repositories. There is currently great interest and scrutiny of existing clinical trial data sharing efforts,[21-24] many of which require submission of a research proposal, as does BioLINCC, and some of which only make data

available via a virtual, secure data sharing environment, as opposed to BioLINCC which provides de-identified data directly to approved researchers. One recently study evaluated how many clinical trials were publicly available to the research community through 3 open access data sharing platforms: ClinicalStudyDataRequest.com, the Yale University Open Data Access (YODA) Project, and the Supporting Open Access for Researchers (SOAR) Initiative, finding that while more than 3000 trials were available, only 15.5% had been requested by a limited number of investigators.[25] The authors concluded that data sharing efforts are being underutilized, implicitly questioning the value of continued resource investment. However, the results of our survey of BioLINCC users suggests this conclusion may be premature, as use of data from these open access platforms can be expected to grow with time, although more remains to ensure the use of these data, and the successful completion and publication of the resulting research, to justify the investments being made in data sharing.

A third limitation of our study is that some information of interest was not asked in order to reduce survey response burden, including questions asking about the time and effort invested to manage and analyze the data from BioLINCC and the impact of the publications resulting from the research project. Finally, our study made no attempt to judge the impact of the research that was able to be completed because of the clinical research data made available through BioLINCC. Other efforts should consider whether the investment being made by NIH and NHLBI in data sharing is justified by the information and knowledge being generated for medical science and society.

In conclusion, we found that the vast majority of investigators who had requested and received access to clinical research data from BioLINCC between 2007 and 2014 had either succeeded in completing their research project or reported being actively involved data analysis or manuscript preparation. In aggregate, more than 100 research projects were completed as a result of respondent investigators using data made available through BioLINCC. Experience with BioLINCC illustrates the

potential of data sharing for the broader scientific field and the importance of funding these efforts, particularly when collecting data of similar size and scope is not feasible for many investigators.

Acknowledgements

The authors would like to acknowledge Mr. Sean Coady of the National Heart, Lung, and Blood Institute (NHLBI) of the National Institutes of Health (NIH), who provided feedback during survey development and assistance with contact information for potential survey respondents, as well as Ms. Tiffany Chang, MPH, Ms. Julia Eichenfield, and Mr. Daniel Shaw, who provided background research and assistance with survey distribution during the course of student employment / voluntary summer research experience at the Yale-New Haven Hospital Center for Outcomes Research and Evaluation (CORE).

Contributors: Study concept and design (Ross, Krumholz, Gross); Acquisition of data (Ritchie, Finn); Analysis and interpretation of data (all authors); Drafting of manuscript (Ross, Ritchie, Finn); Critical revision of manuscript (all authors); Statistical analysis (Ross, Finn); Study supervision (Ross).

Funding/support and role of the sponsor: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. The authors assume full responsibility for the accuracy and completeness of the ideas presented.

Competing Interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and all authors declare (currently or formerly) receiving support through Yale University from Medtronic, Inc. and Johnson and Johnson to develop methods of clinical trial data sharing and from the Blue Cross Blue Shield Association (BCBSA) to better understand medical technology evidence generation. Drs. Krumholz and Ross receive support through Yale University from the Centers of Medicare and Medicaid Services (CMS) to develop and maintain performance measures that are used for public reporting and from the Food and Drug Administration (FDA) to develop methods for post-market surveillance of medical devices. Dr. Krumholz chairs a cardiac scientific advisory board for UnitedHealth. No other disclosures were reported.

Ethical Approval

Ethics approval from the Yale University School of Medicine Human Research Protection Program was obtained prior to study conduct and consent was considered to be implied when participants completed the online survey.

Data sharing

Requests for statistical code and the dataset can be made to the corresponding author at joseph.ross@yale.edu. The dataset will be made available via a publicly accessible repository on publication, at the Dryad Digital Repository (datadryad.org).

Transparency

The lead author (JSR) affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

License

REFERENCES

1. Pharmaceutical Research and Manufacturers of America (PhRMA) and European Federation of Pharmaceutical Industries and Associations (EFPIA). Principles for Responsible Clinical Trial Data Sharing. Our Committment to Patients and Researchers. July 2013. Availabe at http://phrma.org/sites/default/files/pdf/PhRMAPrinciplesForResponsibleClinicalTrialDataSharing.pdf. Last Accessed July 14, 2016.
2. Institute of Medicine of the National Academies of Science. *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risks*. Washington, DC: National Academies Press, 2015.
3. Taichman DB, Backus J, Baethge C, et al. Sharing Clinical Trial Data: A Proposal From the International Committee of Medical Journal Editors. Annals of internal medicine 2016;**164**(7):505-06.
4. Rathi V, Dzara K, Gross CP, et al. Sharing of clinical trial data among trialists: a cross sectional survey. BMJ 2012;**345**:e7570.
5. 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.
6. Savage CJ, Vickers AJ. Empirical study of data sharing by authors publishing in PLoS journals. PloS one 2009;**4**(9):e7078.
7. Tenopir C, Allard S, Douglass K, et al. Data sharing by scientists: practices and perceptions. PloS one 2011;**6**(6):e21101.
8. 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.
9. Oushy MH, Palacios R, Holden AE, et al. To Share or Not to Share? A Survey of Biomedical Researchers in the U.S. Southwest, an Ethnically Diverse Region. PloS one 2015;**10**(9):e0138239.
10. Tudur Smith C, Dwan K, Altman DG, et al. Sharing individual participant data from clinical trials: an opinion survey regarding the establishment of a central repository. PloS one 2014;**9**(5):e97886.

11. Hopkins C, Sydes M, Murray G, et al. UK publicly funded Clinical Trials Units supported a controlled access approach to share individual participant data but highlighted concerns. Journal of clinical epidemiology 2016;**70**:17-25.

12. Giffen CA, Carroll LE, Adams JT, et al. Providing Contemporary Access to Historical Biospecimen Collections: Development of the NHLBI Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC). Biopreservation and biobanking 2015;**13**(4):271-9.

13. Coady SA, Wagner E. Sharing individual level data from observational studies and clinical trials: a perspective from NHLBI. Trials 2013;**14**:201.

14. U.S. National Institutes of Health, National Heart L, and Blood Institute,, Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC). Publications associated with available BioLINCC resources. Availabe at https://biolincc.nhlbi.nih.gov/publications/. Last Accessed January 8, 2016.

15. Hudson KL, Collins FS. Sharing and reporting the results of clinical trials. JAMA : the journal of the American Medical Association 2015;**313**(4):355-6.

16. Fowler FJ. *Survey Research Methods*. Newbury Park, CA: Sage, 1993.

17. Sudman S, Bradburn NM. *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco, CA: Jossey-Bass, 1982.

18. Asch DA, Jedrziewski MK, Christakis NA. Response rates to mail surveys published in medical journals. Journal of clinical epidemiology 1997;**50**(10):1129-36.

19. Keyhani S, Federman A. Doctors on coverage--physicians' views on a new public insurance option and Medicare expansion. The New England journal of medicine 2009;**361**(14):e24.

20. Shanafelt TD, Boone S, Tan L, et al. Burnout and Satisfaction With Work-Life Balance Among US Physicians Relative to the General US Population. Archives of internal medicine 2012:1-9.

21. Longo DL, Drazen JM. Data Sharing. The New England journal of medicine 2016;**374**(3):276-7.

22. Sydes MR, Johnson AL, Meredith SK, et al. Sharing data from clinical trials: the rationale for a controlled access approach. Trials 2015;**16**:104.

23. Tudur Smith C, Hopkins C, Sydes MR, et al. How should individual participant data (IPD) from publicly funded clinical trials be shared? BMC medicine 2015;**13**:298.

24. Strom BL, Buyse M, Hughes J, et al. Data sharing, year 1--access to data from industry-sponsored clinical trials. The New England journal of medicine 2014;**371**(22):2052-4.

25. Navar AM, Pencina MJ, Rymer JA, et al. Use of Open Access Platforms for Clinical Trial Data. JAMA : the journal of the American Medical Association 2016;**315**(12):1283-4.

FIGURE LEGENDS

Figure 1: Inclusion flow chart used to identify potential survey respondents: investigators who had requested and received access to clinical research data from BioLINCC between 2007 and 2014.

Figure 2: Factors influencing decision to request clinical research data through BioLINCC between 2007 and 2014 (n=195).

Note: Respondents were able to select multiple answers in response to this question.

Figure 3. Flow chart showing completion rates of research projects using clinical research data requested from BioLINCC between 2007 and 2014.

Note: Respondents were able to select multiple answers in response to this question.

Table 1. Characteristics of survey respondents and non-respondents.

	Respondents, No. (%) (n=195)	Non-respondents, No. (%) (n=246)	P value
Investigator based in the US?			
Yes	163 (84)	211 (86)	0.53
No	32 (16)	35 (14)	
Investigator based at academic institution?			
Yes	149 (76)	196 (80)	0.41
No	46 (24)	50 (20)	
Investigator based at for-profit institution?			
Yes	5 (3)	4 (2)	0.49
No	190 (97)	242 (98)	
Investigator's total submitted requests to BioLINCC (ever), No.			
1	120 (62)	169 (69)	0.12
> 1 (includes renewals)	75 (38)	77 (31)	
Data sets requested, No.			
1	152 (78)	171 (70)	0.10
2-4	31 (16)	58 (24)	
5-9	7 (4)	14 (6)	

10+	5 (3)	3 (1)	
Request year			
2006	0 (0)	1 (<1)	
2007	13 (7)	17 (7)	
2008	4 (2)	16 (7)	
2009	9 (5)	23 (9)	
2010	6 (3)	17 (7)	0.004
2011	12 (6)	26 (11)	
2012	43 (22)	55 (22)	
2013	47 (24)	39 (16)	
2014	61 (31)	52 (21)	
External funding to support the research project?			
Yes	74 (38)	77 (31)	
No	97 (50)	111 (45)	0.009
Unknown	24 (12)	58 (24)	

Table 2. Sociodemographic and professional characteristics of survey respondents (n=195).

Characteristic	No (%) of respondents
Age	
34 years or younger	29 (15)
35-49 years	97 (50)
50-64 years	47 (24)
65 years or older	14 (7)
Prefer not to answer	8 (4)
Gender	
Male	116 (59)
Female	74 (38)
Prefer not to answer	5 (3)
Race	
White	133 (68)
Asian	35 (18)
Black or African American	10 (5)
Other	3 (2)
Prefer not to answer	14 (7)
Ethnicity	
Hispanic or Latino	8 (4)
Not Hispanic or Latino	175 (90)
Prefer not to answer	12 (6)
Primary employer	

Academic Institution	165 (85)
Non-Profit Organization	14 (7)
Government	8 (4)
Private Industry	4 (2)
Other	4 (2)
Career stage	
In training (< 3 years of active engagement in clinical research, still receiving formative training in research methods)	43 (22)
Early stage career (3-10 years of active engagement in clinical research)	83 (43)
Established in the field (> 10 years of active engagement in clinical research)	69 (35)
Ever been closely involved (as PI or Co-PI) in the conduct of a randomized controlled trial?	
Yes	82 (42)
Ever deposited clinical trial data in the BioLINCC repository?	
Yes	5 (3)

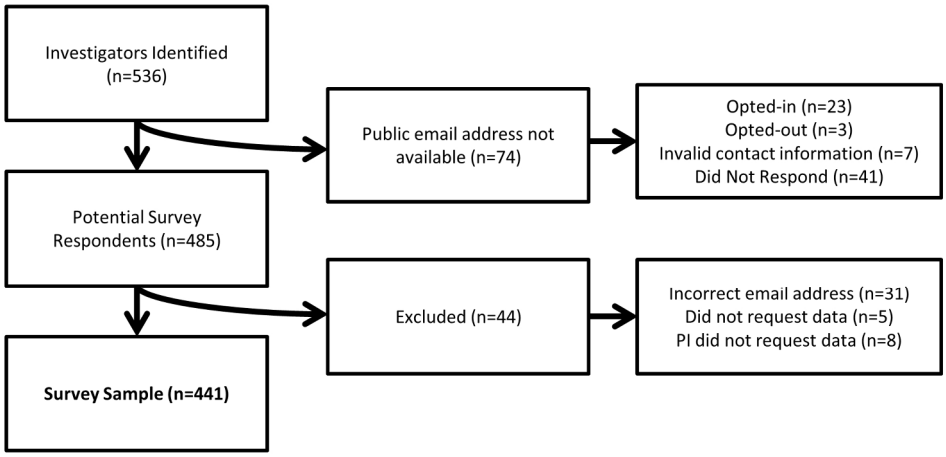

Figure 1: Inclusion flow chart used to identify potential survey respondents: investigators who had requested and received access to clinical research data from BioLINCC between 2007 and 2014.

190x127mm (300 x 300 DPI)

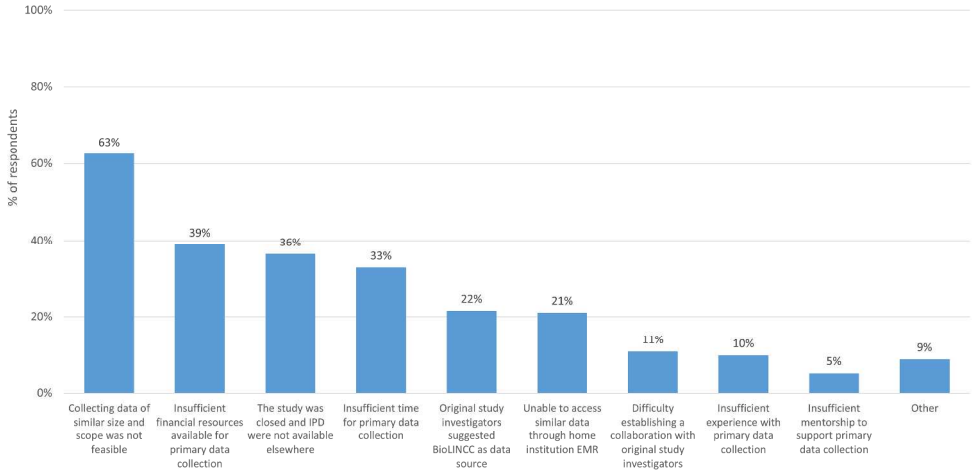

Figure 2: Factors influencing decision to request clinical research data through BioLINCC between 2007 and 2014 (n=195).
Note: Respondents were able to
338x190mm (300 x 300 DPI)

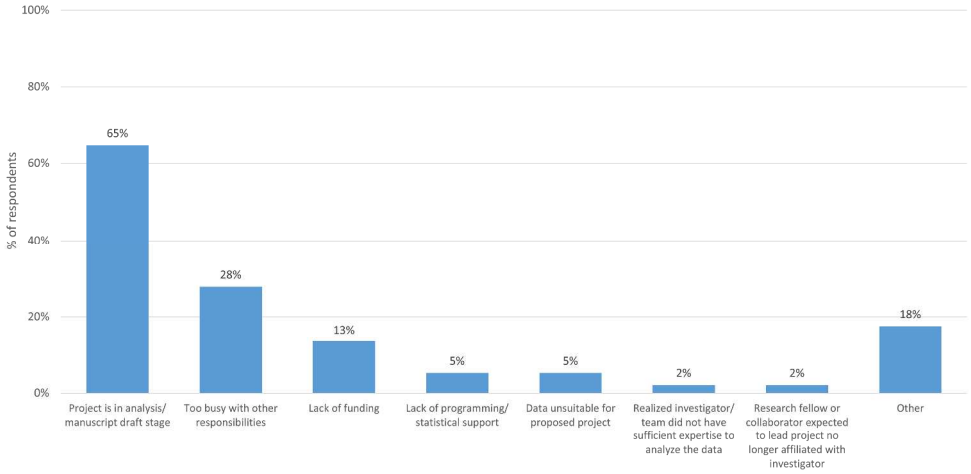

Figure 3. Flow chart showing completion rates of research projects using clinical research data requested from BioLINCC between 2007 and 2014.
Note: Respondents were able to
338x190mm (300 x 300 DPI)

Data Sharing through a NIH Central Database Repository:

A cross-sectional survey of BioLINCC users

APPENDIX

Table of Contents

BioLINCC Opt-in/Opt-out E-mail ..... page 2

Yale University Invitation E-mail ..... page 3

Survey ..... page 4

BioLINCC Opt-in/Opt-out E-mail

In an effort to better understand the benefits, experiences, issues, and barriers to investigators requesting data from data repositories, Yale University will be conducting a short survey of investigators that received data from the NHLBI Data Repository at any time between 2007 and 2014. The survey should take approximately 15 minutes to complete and will provide valuable information on the experiences of users of data repositories. Please note that your responses to the survey will be completely anonymous. NHLBI will not receive nor have any access to any of the survey responses. NHLBI may request specific tables to further the Institute's understanding of potential areas for improvements; however, your individual responses will remain anonymous.

At the conclusion of the survey and publication of findings, your contact information will be permanently removed from all Yale University systems.

Please respond with either a 'Yes' indicating that NHLBI has your permission to share your contact information only with Yale University and only for the purpose of carrying out the survey of NHLBI Data Repository investigators. No permission to share is implied for any other purpose with any other third party.

Or Respond with a 'No' indicating that you do not wish to participate in the survey.

Yale University Invitation E-mail

Subject: Yale Survey on Using Data from NHLBI's Data Repository (BioLINCC)

Yale University, in collaboration with the National Heart, Lung, and Blood Institute (NHLBI) of the NIH, is conducting a short survey of investigators that received data from the NHLBI Data Repository (BioLINCC) at any time between 2007 and 2014. This survey is intended to better understand the benefits, experiences, issues, and barriers to investigators requesting data from this repository.

After having the opportunity to use data from BioLINCC, we hope that you will consider providing feedback on this valuable resource so that efforts can be made to enhance its use by others.

Participation in the survey is voluntary and responses will be anonymized. This survey is expected to take no more than 15 minutes to complete and all participants will be automatically entered in a drawing to win one of five Amazon.com© gift certificates worth $100.

Please complete this survey by June 4, 2015.

Thank you for your participation!

Survey

Yale University is conducting a survey of investigators who have accessed clinical research data through the Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC), at the U.S. National Heart, Lung, and Blood Institute (NHLBI) of the NIH, from 2007 through 2014.

Specifically, we are interested in investigators' experiences with the data and perceptions of the value, importance, and challenges of data sharing initiatives such as this one.

Results from this project are intended to improve investigators' experience with BioLINCC, as well as to inform future clinical data sharing efforts, and are intended to be published in the biomedical literature.

There are no physical risks associated with this project. Participation is voluntary and responses will be anonymized. This survey is expected to take no more than 15 minutes to complete and all participants will be automatically entered in a drawing to win one of five Amazon.com© gift certificates worth $100. Completion of the survey indicates consent to participate.

Please contact us at jessica.ritchie@yale.edu or (203) 200-5346 if you have questions or concerns related to the survey.

I. Reasons for Data Request

1. For what primary research purpose(s) did you request data through NHLBI's Data Repository (BioLINCC)? (Please check all that apply)
   - To conduct an independent scientific study/studies
   - To conduct a pilot/preliminary analysis
   - To conduct an analysis with bio-specimens
   - To learn more about the BioLINCC data request process
   - Other (please specify): [Free text field]

2. Did any of the following influence your decision to request data through NHLBI's Data Repository (BioLINCC)? (Please check all that apply)
   - The study was closed and individual participant-level data were not available elsewhere
   - Original study investigators suggested BioLINCC as the appropriate data source
   - Difficulties in establishing a collaboration with original study investigators
   - Collecting data of similar size and scope was not feasible
   - Insufficient financial resources available for primary data collection
   - Insufficient time for primary data collection
   - Insufficient experience with primary data collection
   - Insufficient mentorship to support primary data collection
   - Unable to access similar data through home institution electronic medical records
   - Other (please specify): [Free text field]

3. How did you intend to use the data? (Please check all that apply) *[**BIFURCATION QUESTION**: If participants answer "To be combined", they will be presented with Question #4. If participants answer "As a standalone data source" or "Other", they will be presented with Question #6.]*
   - As a standalone data source

To be combined with other data sources from BioLINCC
To be combined with other data sources not from BioLINCC (i.e., non-BioLINCC studies, other public data)
Other (please specify): [Free text field]

4. When combining the requested data with other data sources, was the purpose of the project to conduct a meta-analysis? *[**NOTE:** Question #4 only asked of participants who answered "To be combined" to Question #3; **BIFURCATION QUESTION**: If participants answer "Yes", they will be presented with Question #5. If participants answer "No", they will be presented with Question #6.]*
   Yes
   No

5. For the meta-analysis, which of the following was planned? (Please check all that apply) *[**NOTE:** Question #5 only asked of participants who answered "Yes" to Question #4.]*
   Summary-level data meta-analysis
   Participant-level data meta-analysis
   Other (please specify): [Free text field]

6. What was your primary research objective for the data requested through NHLBI's Data Repository (BioLINCC)? (Please check all that apply) [**BIFURCATION QUESTION**: *If participants answer "New research", they will be presented with Question #7. If participants answer "Replication research" they will be presented with Question #8. If participants answer "Other", they will be presented with Question #9.]*
   New research
   Replication research
   Other

7. You indicated that your primary research objective was New research. Which of the following further describe this objective? (Please check all that apply) *[**NOTE:** Question #7 only asked of participants who answered "New research" to Question #6.]*
   To examine secondary endpoints
   To examine subgroup populations
   To leverage the data for a research question unrelated to the original research design (i.e., examine lost-to-follow-up rates or endpoints used in clinical trials)
   To leverage the data to create a cohort for comparison to another study
   Other (please specify): [Free text field]

8. You indicated that your primary research objective was Replication research. Which of the following further describe this objective? (Please check all that apply) *[**NOTE:** Question #8 only asked of participants who answered "Replication research" to Question #6.]*
   Replicate the main study primary endpoint findings
   Replicate the main study secondary endpoint findings
   Replicate the main study subgroup findings (for primary and/or secondary endpoints)

Other (please specify): [Free text field]

9. You indicated that your primary research objective was Other. Which of the following further describe this objective? (Please check all that apply) *[**NOTE:** Question #9 only asked of participants who answered "Other" to Question #6.]*
   Statistical methods research
   Epidemiological research
   Preliminary research to be used as part of a grant proposal
   Other (please specify): [Free text field]

10. Did your research focus on a medical product intervention (i.e., drug, biologic, medical device? *[**BIFURCATION QUESTION**: If participants answer "Yes", they will be presented with Question #11. If participants answer "No", they will be presented with Question #12.]*
   Yes
   No

11. What was the focus of your primary research question? (Please check all that apply) *[**NOTE:** Question #11 only asked of participants who answered "Yes" to Question #10.]*
   Efficacy
   Safety
   Pharmacodynamics
   Other (please specify): [Free text field]

II. Interactions with Original Study Investigators

12. Prior to or after requesting data through NHLBI's Data Repository (BioLINCC), did you contact the original study investigators to obtain the data? *[**BIFURCATION QUESTION**: If participants answer "Yes", they will be presented with Question #13. If participants answer "No", they will be presented with Question #16.]*
   Yes
   No

13. Did the original study investigators approve your data request? [***NOTE:** Question #13 only asked of participants who answered "Yes" to Question #12; **BIFURCATION QUESTION**: If participants answer "Yes", they will be presented with Question #14. If participants answer "No", they will be presented with Question #15.]*
   Yes
   No

14. If the original study investigators approved your data request, for what reason(s) did you also request data through BioLINCC? (Please check all that apply) *[**NOTE:** Question #14 only asked of participants who answered "Yes" to Question #13.]*
   Wanted to validate data made available by original study investigators

Original study investigators required co-authorship to make data available
Original study investigators required control of publication to make data available
Original study investigators required control of study design to make data available
Original study investigators required control of data analysis to make data available
Data made available by original study investigators had no or poor accompanying documentation
Data made available by original study investigators were poorly organized and could not be prepared for analysis
More straightforward to access data through BioLINCC
Other (please specify): [Free text field]

15. What reasons did the original study investigators provide for not approving your data request? (Please check all that apply) *[**NOTE:** Question #15 only asked of participants who answered "No" to Question #13.]*

No interest in collaborating with external investigators
Data cannot be made available to external investigators because original human subject consent forms do not allow
Data cannot be made available to external investigators because of intellectual property issues
Data cannot be made available to external investigators because of patient confidentiality issues
BioLINCC was suggested as the appropriate data source
A reason was not provided
Other (please specify): [Free text field]

16. Prior to or after requesting data through BioLINCC, did you contact the original study investigators to request collaboration? *[**BIFURCATION QUESTION**: If participants answer "Yes", they will be presented with Question #17. If participants answer "No", they will be presented with Question #20.]*

Yes
No

17. For what reason(s) did you request collaboration with the original study investigators? (Please check all that apply) *[**NOTE:** Question #17 only asked of participants who answered "Yes" to Question #16.]*

Needed additional data that were not included in the files provided by BioLINCC
Needed additional statistical expertise due to data complexity
Needed additional content expertise due to study design complexity
Needed additional clinical expertise related to study question
Wanted to work with original study investigators
Other (please specify): [Free text field]

18. Did the original study investigators accept your request for collaboration? *[**NOTE:** Question #18 only asked of participants who answered "Yes" to Question #16; **BIFURCATION QUESTION**: If*

participants answer "No", they will be presented with Question #19. If participants answer "Yes", they will be presented with Question #20.]

Yes
No

19. What reasons did the original study investigators provide for not accepting your request to collaborate? (Please check all that apply) [***NOTE:*** *Question #19 only asked of participants who answered "No" to Question #18.]*

A reason was not provided
Original study investigators were too busy
Original study investigators felt the research question was low priority
Original study investigators did not have funds to support collaboration
Other (please specify): [Free text field]

III. Data Repository Experience

20. How did you learn about the NHLBI's Data Repository (BioLINCC)? (Please check all that apply)

Internet search
Communications with NHLBI
Colleagues/other investigators
Directed to BioLINCC by study investigators
Other (please specify): [Free text field]

21. Is there anything you would have liked to have known prior to accessing the BioLINCC data?

No
Yes (please provide further details): [Free text field]

22. Were the data you received from BioLINCC suitable for your originally proposed project? [***BIFURCATION QUESTION****: If participants answer "Yes", they will be presented with Question #24. If participants answer "No", they will be presented with Question #23.*]

Yes
No

23. For what reason(s) was the data unsuitable for your proposed project? (Please check all that apply) [***NOTE:*** *Question #23 only asked of participants who answered "No" to Question #21.*]

Data had no or poor accompanying documentation; could not determine if data were suitable
Data were too complicated to use; could not determine if data were suitable
Data were poorly organized; could not be adequately prepared for analysis
Data had too many missing values; could not be adequately prepared for analysis
Proposed main outcome variable was not available in data
Main outcome variable was not available in data at the time points proposed for study
Proposed main independent variable was not available in data
Other (please specify): [Free text field]

24. Please consider your experience using the data you received from BioLINCC. Do you agree or disagree with the following statement: The documentation and data dictionaries (i.e., meta-data) received from BioLINCC were useful. [***BIFURCATION QUESTION:*** *If participants answer "Somewhat disagree" or "Strongly disagree", they will be presented with Question #25. Otherwise, they will be presented with Question #26.]*
    Strongly agree
    Somewhat agree
    Somewhat disagree
    Strongly disagree

25. You chose "Somewhat disagree" or "Strongly disagree." Please briefly explain your answer. [***NOTE:*** *Question #25 only asked of participants who answered "Somewhat disagree" or "Strongly disagree" to Question #24.]*
    [Free text field]

III. Project Details

26. Has your project been completed? *[**BIFURCATION QUESTION:*** If participants answer "No", they will be presented with Question #27. If participants answer "Yes", they will be presented with Question #31.]
    Yes
    No

27. For what reason(s) was the project not completed? (Please check all that apply) *[**NOTE:** Question #27 only asked of participants who answered "No" to Question #26.]*
    Project is in analysis/manuscript draft stage
    Data unsuitable for proposed project
    Realized investigator/team did not have sufficient expertise to analyze the data
    Lack of funding
    Lack of programming/statistical support
    Too busy with other responsibilities
    Research fellow or collaborator expected to lead project no longer affiliated with investigator
    Investigator no longer active in clinical research
    Others published same/similar work on same/similar data
    Other (please specify): [Free text field]

28. Do you plan to complete the project? *[**NOTE:** Question #28 only asked of participants who answered "No" to Question #26. **BIFURCATION QUESTION:*** If participants answer "No", they will be presented with Question #29. If participants answer "Yes", they will be presented with Question #31.]
    Yes
    No

29. Was there an issue with the BioLINCC data that prevented you from completing the project? *[**NOTE:** Question #29 only asked of participants who answered "No" to Question #28.* ***BIFURCATION QUESTION:*** If participants answer "Yes", they will be presented with Question #30. If participants answer "No", they will be presented with Question #39.]
    Yes
    No

30. Please briefly explain your answer. *[**NOTE:** Question #30 only asked of participants who answered "Yes" to Question #29.]*
    [Free text field]

31. Does your completed or anticipated final project differ from your pre-specified project? *[**NOTE:** Question #31 only asked of participants who answered "Yes" to either Question #26 or #28*. ***BIFURCATION QUESTION:*** If participants answer "Yes", they will be presented with Question #32. If participants answer "No", they will be presented with Question #33. **NOTE:** If participants answer "No" to #26 and "No" to #31, they will be presented with #38]
    Yes
    No

32. In what ways does your completed or anticipated final project differ from your pre-specified project? (Please check all that apply) *[**NOTE:** Question #32 only asked of participants who answered "Yes" to Question #31.]*
    Modified planned data source by combining data received from BioLINCC with other data sources
    Modified planned data source by not combining data received from BioLINCC with other data sources
    Modified study sample
    Modified primary endpoints
    Modified secondary endpoints
    Modified selection of main independent variables
    Modified statistical analysis plan
    Other (please specify): [Free text field]

33. Was your research project published? [***NOTE:** Question #33 only asked of participants who answered "Yes" to Question #26*. ***BIFURCATION QUESTION:*** If participants answer "Yes", they will be presented with Question #34. If participants answer "No", they will be presented with Question #36.]
    Yes
    No

34. In what format was your research project published? (Please check all that apply) [***NOTE:** Question #34 only asked of participants who answered "Yes" to Question #33*.]
    Original research article in a peer-reviewed biomedical journal
    Systematic review/meta-analysis in a peer-reviewed biomedical journal
    Non-systematic review article in a peer-reviewed biomedical journal
    Commentary / viewpoint / editorial in a peer-reviewed biomedical journal
    Letter in correspondence in a peer-reviewed biomedical journal
    Weblog post or other on-line forum

Self-published
Other (please specify): [Free text field]

35. Please provide the publication citation and PubMed ID (if applicable) or other citation (such as web address). [***NOTE***: *Question #35 only asked of participants who answered "Yes" to Question #33.*]
    [Free text field]

36. When attempting to publish your research, were any of the following concerns raised by editors or peer reviewers during the peer-review process? (Please check all that apply)
    I did not attempt to publish
    No substantive concerns were raised beyond minor comments and suggestions to clarify/improve the research
    Concern that I was not one of the original study investigators
    Concern about the original study design that I could not address
    Concern about my research project design that I could not address without additional data
    Concern about my research project design unrelated to the data
    Concern about my research methodology and analysis
    Concern about the importance of my research
    Other (please specify): [Free text field]

37. How many research projects did you complete, or do you plan to complete, through your single request? [***NOTE:*** *Question #38 only asked of participants who answered "Yes" to either Question #26 or #28.*]
    One
    Two
    Three
    Four or more

38. Did using data from BioLINCC aid in any future grant applications? (Please check all that apply)
    Yes, use of the BioLINCC data established a publication record that was then included in a grant application
    Yes, use of the BioLINCC data furthered the understanding of questions of particular interest/identified gaps that then served as the basis of a grant application
    Yes, other (please specify): [Free text field]
    No

39. What was the primary funding source used to support this project?
    Self-funded
    NIH
    Non-NIH Federal
    Non-Profit Organization or Foundation in US
    Industry
    Non-US Government or Organization
    Other (please specify): [Free text field]

40. What additional funding source(s) were used to support this project? (Please check all that apply)
    - Self-funded
    - NIH
    - Non-NIH Federal
    - Non-Profit Organization or Foundation in US
    - Industry
    - Non-US Government or Organization
    - Other (please specify): [Free text field]
    - None

IV. Requestor Demographics

41. Which of the following best classifies your primary employer at the time when you requested data through BioLINCC?
    - Academic Institution
    - Private Industry
    - Non-Profit Organization
    - For-Profit Hospital
    - Government
    - Other (please specify): [Free text field]

42. How would you classify your career status with respect to clinical or epidemiological research at the time when you requested data through BioLINCC?
    - In training (< 3 years of active engagement in clinical research, still receiving formative training in research methods)
    - Early stage career (3-10 years of active engagement in clinical research)
    - Established in the field (> 10 years of active engagement in clinical research)

43. Have you ever been closely involved (as Principal or Co-Investigator) in the conduct of a randomized controlled trial?
    - Yes
    - No

44. Have you ever deposited clinical trial data in the BioLINCC repository?
    - Yes
    - No

45. Please indicate your age range.
    - 34 years or younger
    - 35-49 years
    - 50-64 years
    - 65 years or older
    - Prefer not to answer

46. Please indicate your gender.
    - Male
    - Female

Prefer not to answer

47. Please indicate your ethnicity.
    Hispanic or Latino
    Not Hispanic or Latino
    Prefer not to answer

48. Please indicate your race.
    American Indian or Alaska Native
    Asian
    Black or African American
    Native Hawaiian or Other Pacific Islander
    White
    Other (please specify): [Free text field]
    Prefer not to answer

49. Thank you for completing this survey. To be eligible for entry into the Amazon.com gift certificate drawing, please provide your email address. Your email address will be kept separate from your survey responses in order to ensure anonymity.

50. Would you like to receive a notification when the results of this study are published? If so, please re-enter your email address. Your email address will be kept separate from all other responses to ensure confidentiality.

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	3-4
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	5-6
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	6-7
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6-7
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	6-7
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	8-9
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	8-9
Bias	9	Describe any efforts to address potential sources of bias	9-10
Study size	10	Explain how the study size was arrived at	9-10
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	8-10
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	9-10
		(*b*) Describe any methods used to examine subgroups and interactions	n/a
		(*c*) Explain how missing data were addressed	n/a
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	n/a
		(*e*) Describe any sensitivity analyses	n/a
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	10
		(b) Give reasons for non-participation at each stage	10
		(c) Consider use of a flow diagram	Figure 1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	10, 25-28
		(b) Indicate number of participants with missing data for each variable of interest	10, 27-28
Outcome data	15*	Report numbers of outcome events or summary measures	11-13
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	11-13
		(*b*) Report category boundaries when continuous variables were categorized	n/a
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	n/a
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	n/a
Discussion			
Key results	18	Summarise key results with reference to study objectives	13-15
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	16-17
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	13-17
Generalisability	21	Discuss the generalisability (external validity) of the study results	13-15
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	19

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.