



Published in final edited form as:

Clin Trials. 2013 ; 10(6): 977–986. doi:10.1177/1740774513503522.

The National Drug Abuse Treatment Clinical Trials Network Data Share Project: Website Design, Usage, Challenges and Future Directions

Dikla Shmueli-Blumberg, PhD¹, Lian Hu, PhD, MPH¹, Colleen Allen, MPH¹, Michael Frasketi, BS¹, Li-Tzy Wu, ScD, RN², and Paul VanVeldhuisen, PhD¹

¹The EMMES Corporation, Rockville, MD, USA

²Department of Psychiatry and Behavioral Sciences, Duke University, Durham, NC, USA

Abstract

Background—There are many benefits of data sharing, including the promotion of new research from effective use of existing data, replication of findings through re-analysis of pooled data files, meta-analysis using individual patient data, and reinforcement of open scientific inquiry. A randomized controlled trial is considered as the “gold standard” for establishing treatment effectiveness, but clinical trial research is very costly and sharing data is an opportunity to expand the investment of the clinical trial beyond its original goals at minimal costs.

Purpose—We describe the goals, developments, and usage of the Data Share website (www.ctndatashare.org) for the National Drug Abuse Treatment Clinical Trials Network (CTN) in the US, including lessons learned, limitations and major revisions and considerations for future directions to improve data sharing.

Methods—Data management and programming procedures were conducted to produce uniform and Health Insurance Portability and Accountability Act (HIPAA)-compliant de-identified research data files from the completed trials of the CTN for archiving, managing, and sharing on the Data Share website.

Results—Since its inception in 2006 and through October 2012, nearly 1700 downloads from 27 clinical trials have been accessed from the Data Share website, with the use increasing over the years. Individuals from 31 countries have downloaded data from the website, and there have been at least 13 publications derived from analyzing data through the public Data Share website.

Limitations—Minimal control over data requests and usage has resulted in little information and lack of control regarding how the data from the website are used. Lack of uniformity in data elements collected across CTN trials has limited cross-study analyses.

Conclusions—The Data Share website offers researchers easy access to deidentified data files with the goal to promote additional research and identify new findings from completed CTN studies. To maximize the utility of the website, on-going collaborative efforts are needed to

Corresponding author: Dikla Shmueli-Blumberg NIDA CTN Clinical Coordinating Center The EMMES Corporation 401 N. Washington St., Suite 700 Rockville, MD, 20850 dblumberg@emmes.com.

Declaration of Conflicting Interests: None.

standardize the core measures used for data collection in the CTN studies with the goal to increase their comparability and to facilitate the ability to pool data files for cross-study analyses.

BACKGROUND

Data sharing involves making participant-level research data files available to the broader scientific community. Examples of research data that typically are involved in data sharing include survey or clinical trial data files and other experimental data that serve to support research analysis and findings. In certain fields, research data may include physical collections and bio-specimen samples (e.g., serum, plasma, DNA, and tissue).¹ Data sharing does not refer to summary statistics, drafts of scientific papers, internal communications, laboratory notebooks, or similar items.²

Continued advancements in cyber infrastructure allow for more efficient data acquisition, storage, management and integration, and provide new data sharing opportunities. Particularly in the field of public health, data sharing promotes the translation of research findings into knowledge and practices that can further improve human health conditions.^{3,4} The benefits of data sharing are substantial and include the promotion of new research from effective use of existing data,^{4,7} replication of findings through re-analysis of pooled data files,⁷⁻⁹ meta-analysis using individual patient data, reinforcement of open scientific inquiry,^{6,10,11} and encouragement to develop different theoretical perspectives, especially in an interdisciplinary setting.^{8,10,12} Data sharing adds value with little cost and thus optimizes the use of monetary and time resources,^{4,8} lessens the requirement to recruit and involve individuals in research studies since fewer studies can potentially answer a greater number of research questions¹² and reduces the possibility that funding organizations unknowingly will “double-fund” the same project.⁷

Numerous scientific organizations, journal editors, and research funding agencies advocate sharing data as a vital part of the scientific process.¹³ Among them, the National Institutes of Health (NIH) encourage data sharing with the expectation that it will lead to products and knowledge that will benefit the public.² Clinical trials or studies have long been recognized as an important but costly area of research where sharing data is particularly advantageous, and sometimes vital, in order to reach scientific goals.¹⁴ The high cost of multi-site trials makes it especially important to share data for further secondary data analysis.

The National Institute on Drug Abuse (NIDA) National Drug Abuse Treatment Clinical Trials Network (CTN) is the first and the largest national network of addiction treatment researchers and community treatment programs (CTPs) working cooperatively to conduct multi-site studies of behavioral, pharmacological, and integrated treatment interventions in community-based treatment settings, with the goal of bridging between clinical research and practice.¹⁵ In particular, the objective is to translate findings from drug abuse treatment trials into knowledge that may be applied in the real-world treatment settings to improve treatment practices for diverse patient populations and to inform the transfer of research results to medical clinicians, other health professionals and patients. The Data Share website (www.ctndatashare.org) presents an ideal platform for sharing the wealth of research data

from completed CTN treatment trials, and allows for broad dissemination of study data and, ultimately, greater impact of these drug treatment studies on the addiction treatment field.

We describe the goals and development of the CTN Data Share website, from its initial inception in 2006 through its current form. From the experiences gathered throughout these years, we share and explore various lessons learned, limitations as well as major revisions of the website, and considerations for future directions for data sharing, in particular as it relates to the NIDA CTN, but with applicability to other data sharing initiatives.

METHODS

CTN Database Overview and Design

The CTN Data Share website was designed with several goals in mind. First, the website allows for both archiving and managing de-identified research data from completed network trials in a standard and transparent manner. Making trial data easily accessible on the Data Share website serves to increase data use and thereby to increase scientific productivity and to optimize the use of resources originally invested in the trials. The public database provides numerous opportunities for interested researchers to apply different analytic plans to stored data regarding safety and effectiveness of various interventions based on a single study or multiple studies. It also offers an opportunity for professors to use the de-identified data in their classrooms for learning and practice purposes. The Data Share website, which is linked to the larger CTN website, increases the visibility of the CTN, and attracts international collaborators as well. For example, several participants from the NIDA International Program, the INVEST Drug Abuse Research fellowship, have come to the United States to learn about clinical trials and have completed their fellowships under the mentorship of CTN investigators. These participants and the international research communities in general, are able to use the public data to learn more about the CTN research and designs, as well as to use the collected data for additional analysis and preparation of manuscripts for publication. The CTN Data Share website also promotes analysis of pooled data files to expand knowledge about demographic and clinical characteristics of drug users who participated in the CTN studies, evaluate the quality and utility of diagnostic and clinical instruments used by multiple trials, facilitate the testing of newer hypotheses or subgroup differences in treatment responses, and generate opportunities for national and international collaboration.¹⁶⁻²² Furthermore, specific advantages of analyzing cross-study data files are to increase statistical power by pooling data and, potentially, to generalize study findings to substance use subpopulations, such as minorities.²³ Finally, the database enables the translation of research results into knowledge, products and practice to improve public health¹⁴, one of the fundamental missions of the NIH.

Preparation of Data for Sharing

To increase timely use of data for secondary data analysis research from closed trials, the CTN's Data and Statistics Center has worked to produce the de-identified data files and release the data files and related study information to the CTN Data Share website in a timely manner, i.e., either 18 months after the completion of a trial or after the primary

manuscript has been accepted for publication, whichever comes first. As of July 2012, the CTN Data Share contained data files from 27 completed clinical trials (**Table 1**).

Prior to posting to the website, trial data are de-identified to maintain study participant privacy, consistent with requirements of both the Health Insurance Portability and Accountability Act (HIPAA) and the Federal Policy for the Protection of Human Subjects (Title 45 CFR part 46, Human Subjects Protection). De-identified data sets for a study are prepared by staff who first critically review case report forms and the accompanying data dictionary to identify a list of variables that contain, or potentially could contain, identifying information, such as personal identifiers (e.g., names, initials), dates, free text fields, or any information that alone or in combination with other information potentially could identify a study participant. The variables then are removed or modified in a copy of the raw dataset, resulting in a de-identified dataset. To illustrate, all dates are replaced with days on study, where the anchor date is date of randomization; date of birth is replaced with age at randomization; site identifiers are removed; and participant study identifiers are replaced with a randomly-generated numeric identifier.

De-identification of the data is a necessary step to ensure protection of participant information that is available to the public. Most data that are removed or recoded during de-identification either are not relevant to analyses so that the de-identified data remain useful for analyses; little information is lost by de-identification. For example, dates converted to days on study allow time-to-event analyses. One exception is that with individual site identifiers being removed, rather than recoded, analyses adjusting for or investigating site effects are not possible using data from the CTN Data Share website.

The CTN Data Share website is organized by protocol or study; the material specific to each study is added to the website as it becomes available. For each study, a number of important study -specific documents are also available to users of the website; are accessed separately from the data files and therefore do not require registration to download them. For each trial on the Data Share website, the following information can be downloaded:

- Data files intended to allow researchers to perform statistical analyses are available in two formats: comma separated variable files (American Standard Code for Information Interchange, ASCII) and Statistical Analysis System format (SAS) transport files;
- Descriptive metadata such as data dictionaries and annotated case report forms are provided for the data files, as well as overall descriptive data , such as study number, title, and a summary of the design;
- Study protocol, a document that describes the scientific rationale, objectives, design, methodology, planned analyses, and organization of the study;
- De-identification notes, which describe in detail the study-specific de-identification process;
- Primary manuscript reference, when available; and

- [Website links](#) external to Data Share integrated into the website to reference study summaries provided on the CTN website and the study description on clinicaltrials.gov.

In addition, taxonomy fields are provided for study keyword, investigator and assessments associated with the study to aid in searching the Data Share website. This taxonomy, a central feature of the Data Share website, creates relationships between the various data points and links content across studies using vocabulary categorization. This classification system links data in such a way that users can find content (e.g., studies) they may not have been aware existed or were applicable to their search. For example, the first CTN study, CTN-0001 (“Buprenorphine/Naloxone versus Clonidine for Inpatient Opiate Detoxification”), was assigned the following keywords: clonidine, buprenorphine, opiate, and naloxone. The Principal Investigator (PI) is listed for the study, as well as the seven assessments used (e.g., the Addiction Severity Index Lite). Using these entries, it is possible to see on how many other CTN studies were associated with the PI, which other studies use assessments similar to those used in CTN-0001, and which studies have been assigned similar keywords. All site content is indexed and a robust search engine is available for users to explore the content in depth.

Administrators of the website, enter data (e.g., study title, study number) directly into the website by way of data input forms with fields for each data point. Controls ensure consistency throughout the site and prevent duplication of data or minor differences such as typographical errors. A release date is listed for each study to allow for chronological sorting and filtering.

Support is provided by the Data and Statistics Center by public users who send comments and questions though a link on the Data Share website. Both technical and research-oriented questions have been received; the volume of questions has been manageable so far.

Expansion of the Original Database

In 2011, a new feature was integrated into the CTN Data Share website to provide detailed and categorized assessment information for all studies, which allows public users to view and browse all study assessments used per study as well as identify studies that used similar assessments. This feature facilitates the identification of similar studies for pooled analyses. A listing of all study assessments can be displayed and filtered according to defined categories to display a desired subset of assessments. In addition to basic identification fields (description and abbreviation), some assessments provide an external link to 1 other descriptive information at the Alcohol and Drug Abuse Institute (ADAI) Library hosted on the University of Washington's website. On the CTN Data Share website, the relationships between study assessments and protocol are established, with each study assessment classified under one of nine general categories and associated subcategories as necessary (**Table 2**). The two categories with the largest number of assessments connected with NIDA CTN trials are Substance Use and Mental Health, which have been further partitioned into subcategories, e.g., the subcategory Depression within the category of Mental Health. The full assessment list also can be filtered by category to allow displays of only those assessments that fall within a specified category, e.g., Mental Health. A decision was made

to defer the categorization of assessments related to biological/physical examinations, such as vital sign and pregnancy test results, and study management data, such as protocol violations and medication logs in this first phase of the project. One of the goals for future enhancement of the website is to expand the search criteria to include the latter types of assessments.

A query tool was created to allow researchers to search for specific assessments across studies, for example, to ascertain how many studies have used the Life Events Checklist, or to compare the assessments used in two or more studies, for example, to determine whether studies CTN-0001 and CTN-0019 used any of the same measures. Because the assessments are categorized into higher level substantive categories, it is also possible to search for a certain construct rather than a specific measure. For example, rather than just searching for studies that have used the Life Events Checklist, researchers can search for studies that have used any Post Traumatic Stress Disorder (PTSD) scale. The results of the query then produces a list of studies that used similar or identical measures (e.g., Post Traumatic Stress Disorder Symptom Scale-Self Report). Furthermore, it is possible to choose any combination of measures across different categories. For example, a researcher interested in treatment of tobacco dependence with co-occurring psychiatric disorders can choose a variety of Tobacco and Mental Health assessments to see which studies have used any or all such assessments.

Layout, Technical Design and Cost

To power the efficiency of search, a popular open-source Content Management System (CMS) was utilized and built on a standard Linux, Apache, MySQL and PHP (LAMP) stack within a virtual and fully redundant data center. This feature provides for a powerful theming and search engine, content management, menu system and access control, with minimal programming effort. Several custom modules also have been developed to support the assessment query and discovery features.

Each element within the website is entered via a user interface provided to appropriately privileged and trained users. Each data point is entered into a discrete form element, which ensures proper data entries and enforces field requirements (date, number, text). When necessary, additional description or help information is provided to the user to ensure appropriate data entry. To guarantee a clean and predictable appearance and site layout, no HTML or complex code is entered by the user. All page formatting is controlled with templates that ensure consistency and maintain a familiar look and feel to each page. Templates also facilitate the data entry process, so that minimal training is necessary for content entry.

The website was developed at relatively low cost, in part due to the use of the open-source CMS. Layout and theming efforts as well as some functional programming accounted for some of the cost. A fair amount of the development effort was incurred by project staff (non-developers) to enter the study information, links, documents and supporting meta-data. Similarly, maintaining the website over time, including tasks such as the addition of studies, implementation of a query tool, and responding to users have relatively low cost. The major

expenses in the implementation of the website were associated with converting the data to the Study Data Tabulation Model (SDTM) format.

CTN Data Share Usage

One of the goals prompting the development of the Data Share was effective dissemination of results to the larger research community in an easy and user-friendly manner. To achieve this goal, relevant study design documents, e.g., the protocol, annotated case report forms, have been made available to all public users. To understand the extent to which CTN data are downloaded (or used) by research or academic communities, potential data users are required to provide basic contact information, including the name of the investigator, his/her position, affiliation, email address, and country, before de-identified data files can be downloaded. This set of information is kept on file but it has not been verified.

Prior to downloading the data, the potential recipient of the data is asked to review and agree to the terms and conditions for use of the data: 1) to agree not to attempt to establish the identity of any study participant; 2) to retain control of the data and not to transfer it to other entities; 3) to obtain Institutional Review Board approval of the planned research, when applicable; 4) to acknowledge the CTN in any publications or presentations; 5) to maintain security and privacy of the data; 6) to inform the NIDA Center for Clinical Trials Network (CCTN) when the research that uses the data is published; and 7) to agree that the CCTN may contact the recipient regarding the use of the data. Whenever the potential recipient of the data agrees to the terms and conditions, the applicant is asked to indicate the preferred file format. Upon completion of the web form, links to downloadable zipped files that contain the data, including study title and URL information, are made available immediately, in either ASCII or SAS format.

Another goal of the Data Share website is to encourage further analyses and promote new research. Currently, the CTN Data Share contains data files from 27 completed studies (**Table 1**). Of the 27 studies, 24 were randomized controlled trials (RCTs), 3 were surveys among clinicians and other workforces of Community Treatment Programs (CTPs) within the NIDA CTN. Various interventions were examined in the RCTs, of which 5 were pharmacological, 15 were psychosocial/behavioral and 4 were combined pharmacological and psychosocial/behavioral interventions.

RESULTS

Over a three-year period (2010-2012), 20 requests for information were received; most took minimal effort in response. Since its inception in 2006 and through October 2012, there have been nearly 1700 downloads from the Data Share website (**Figure 1**); use has increased over the years. Individuals from 31 countries have downloaded data thus far, with the most downloads for researchers in the United States, India, and China. The number of requests for data from each specific study generally corresponds with the length of time the study data have been publically available. Thus, the most frequently requested data set is for the first CTN study, Buprenorphine/Naloxone versus Clonidine for Inpatient Opiate Detoxification. To our knowledge, there are at least 13 publications in the English language derived from analyzing data through the Data Share website. Some of these involve secondary analyses of

individual studies,^{21,24-26} while others have merged data from several CTN studies.^{22,27-30} For example, Wu et al. applied factor and item response theory analyses to evaluate psychometric information and the quality of diagnostic tools in CTN studies.^{16,17,21,22,26} The results support diagnostic assessments used by the CTN and generate empirical data about the classification of DSM-IV substance use disorders to inform DSM5. Lindblad et al. examined the safety data (adverse events and serious adverse events) from 17 completed CTN studies and developed a tailored safety strategy to reduce reporting burden of irrelevant safety events in clinical trials.²⁹ Additional research findings are anticipated to be published in the near future through secondary use of Data Share data.

DISCUSSION

The many findings from secondary analyses of CTN data files not only have generated timely empirical data to inform the ongoing design and conduct of addiction treatment studies and analyses of data from minorities and women to address NIH missions, but also have helped to reveal comorbidities and substance abuse patterns of CTN participants. For example, most CTN studies require the use of diagnostic tools to apply inclusion and exclusion criteria for selection of participants. Thus, the quality of a diagnostic tool is critical to valid interpretation of study findings.^{16,17,21,22,26} Women and members of minority groups often are underrepresented in clinical trials; the NIH requires the inclusion of these populations in clinical studies unless there are appropriate reasons for excluding them from the research. The availability of data files from multiple CTN studies not only allows investigators to examine women and minority groups' ongoing participation and retention in CTN studies,³¹ but also helps elucidate gender and racial/ethnic issues in clinical measures, subgroup analysis, polydrug use patterns, comorbidity, and HIV risk behaviors.^{19,23,27,28,30,32,33} These findings provide invaluable information to the addiction treatment field and inform further refinement of newer study designs.^{23,34} While the costs of maintaining the website are fairly low, the potential benefits of the use of the data are high, and continue to grow as the existing studies continue to be used and data from new studies are added to the website.

Challenges for the Management of the CTN Data Share Data

Making clinical research data accessible to researchers has numerous advantages. Data sharing ensures that tax payers' investment in health research has the potential to yield maximum knowledge and health benefits. The complexity of CTN Data Share lies in implementing, monitoring, and supporting existing and newer data files from ongoing studies.

One of the challenges in managing and using the Data Share website involves the minimal control exerted over data requests and usage. The initial goal for the Data Share website was to make it user-friendly and accessible to encourage public usage. The procedure for requesting data files is simple, free, and quick. The requested data files are delivered immediately upon submission of a simple request form and acceptance of the terms of use. Other similar public databases may require a data requester to submit a research protocol, to obtain approval from the Institutional Review Board with which the requester is affiliated, to

pay fees for the use of data, and to wait for approval.^{35,36} The disadvantage of full and easy public access is that there is little information and a lack of control regarding how the data are used; rarely are notifications received from users who have published findings utilizing the CTN data. We do not know the level of compliance with the website user agreement because compliance with the terms is not monitored or enforced. Although we do not wish to impose measures that would discourage use of the data, one initiative currently under review is to contact the researchers who have requested data and to inquire about their compliance with the terms of the user agreement; for example, whether they have used CTN data in publications or presentations.

Another challenge in the management and use of the Data Share is related to using the legacy-converted SDTM data sets. Since 2004, the FDA has encouraged pharmaceutical industries to submit clinical trial safety data in SDTM format developed by Clinical Data Interchange Standards Consortium (CDISC, <http://www.cdisc.org/>). FDA's intention is to ensure that publically available data standards are in place. These well-defined standards are considered to support multiple uses of the data such as regulatory review, analyses across clinical research data and exchanging data between clinical studies and electronic healthcare records, among other uses (accessed through <http://www.fda.gov/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/ucm269946.htm>). As original CTN data were not collected with case report forms conforming to the CDISC Clinical Data Acquisition Standards Harmonization (CDASH) recommended standards, it has been difficult to ensure the accuracy of mapping from non-standard CTN data into the SDTM format. Technical support questions received by the Data and Statistics Center and NIDA CCTN from Data Share users suggest that it is difficult to extract desired information from legacy converted SDTM data sets. Problems associated with “legacy data conversion” have downgraded the re-usability of Data Share and have required the CTN's Data and Statistics Center to provide additional support for the data users. Subsequently, the FDA also found that the SDTM format data available through “legacy data conversion” had significantly delayed, rather than expedited, the FDA review process as anticipated.³⁷ Consequently, the FDA now requires sponsors to submit both SDTM and analysis data sets to ensure efficient review processes.³⁷

Future Directions and Next Steps for CTN Study Data Collection and Management

Underlying the challenges is the need to establish data collection standards prospectively, i.e., to develop standard measures and case report forms for data collection, in order to avoid data conversion. Standardization ensures comparability and facilitates the ability to pool data files and to conduct cross-study analyses. Although it would be impractical to require every CTN study to use uniform case report forms to collect all data, it is critical, from the public sharing point of view, to identify a core set of data elements for all CTN clinical trials and to standardize the case report forms for these core data elements. The NIDA CTN is striving to streamline clinical research data collection and standards through centralizing the design of their electronic case report forms to enhance data standardization in future studies.

In another large-scale effort to promote standardization of Substance Use Disorders clinical research data and the harmonization of such data, the consensus measures PhenX (Phenotypes and eXposures) project was launched, initially by the National Human Genome Research Institute (NHGRI) and the Office of Behavioral and Social Sciences Research (OBSSR) (<https://www.phenxtoolkit.org/index.php?pageLink=home.more>) but now has involved most institutes at the National Institutes of Health (<https://www.phenx.org/Default.aspx?tabid=235>). PhenX contains a core set of clinical measures for complex diseases, phenotypic traits and environmental exposures across various research fields. Members of NIDA CCTN were involved in the Substance Abuse and Addiction working group that developed PhenX measures related to substance use. These consensus-based measures are free and available to researchers in a web-based toolkit (<https://www.phenxtoolkit.org>). The anticipated widespread adoption of these open-source PhenX measures across research areas should greatly improve the statistical power to detect genotypic-phenotypic associations and facilitate secondary analysis research across different fields. NIDA *strongly encourages* its applicants to incorporate three sets of PhenX data measures into their clinical studies: Core Tier 1 (demographics, tobacco, alcohol and substance use questions), Core Tier 2 (socioeconomic status), and Specialty Areas (e.g., substance use-related co-morbidities and health-related outcomes).³⁸ For future CTN studies, grantees will be required to incorporate Core Tier 1 PhenX measures into case report forms. Implementation of PhenX measures is anticipated to improve standardization and streamlining of data collection in future CTN studies in an effort to improve data re-usability, reduce the cost of research, and, ultimately, benefit people suffering from substance use disorders.

In addition to standardization, other important factors that would enhance the usefulness of the Data Share include ensuring the measurement validity of the assessments used in the studies and recognition that although randomized controlled trials (RCTs) are the gold standard for clinical trials, they are not flawless; other types of study data are valuable as well.

Conclusion

The CTN Data Share website was developed to promote effective use of already collected data to address research questions and to enable analyses that may yield additional clinically important information to inform and improve the quality of drug abuse treatment by making the data from completed addiction treatment studies easily accessible to the broader scientific community. Examination of global data usage suggests that the website is beginning to accomplish these objectives and that the overall usage is increasing over time. Ongoing efforts to standardize the core measures used for data collection in the CTN studies and the incorporation of Tier 1 PhenX measures into the case report forms for upcoming clinical trials should strengthen interoperability and harmonization even further and increase the usefulness of this valuable data sharing website.

Acknowledgments

The work for the National Drug Abuse Treatment Clinical Trials Network (CTN) Data Share website is made possible by the U.S. National Institute on Drug Abuse (NIDA) of the National Institutes of Health (grant number

HHSN271200900034C to The EMMES Corporation, and grant number HSN271200522071C to Duke University). Li-Tzy Wu has received research funding from NIDA (R33DA027503, R01DA019623). The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

References

1. National Science Foundation (NSF). [September 19, 2012] National Science Foundation (NSF) Grant General Conditions (GC-1) Effective. Feb 1. 2012 No.GC-1 (02/12). 2011; http://www.nsf.gov/publications/pub_summ.jsp?ods_key=gc0212.
2. National Institutes of Health Office of Extramural Research. [July 11, 2012] NIH Data Sharing Policy and Implementation Guidance.. Policy&Guidance. 2003. http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm.
3. Walport M, Brest P. Sharing research data to improve public health. *Lancet*. Feb 12; 2011 377(9765):537–539. [PubMed: 21216456]
4. Pisani E, AbouZahr C. Sharing health data: good intentions are not enough. *Bulletin of the World Health Organization*. Jun; 2010 88(6):462–466. [PubMed: 20539861]
5. Fienberg, SE.; Martin, ME.; Straf, ML. *Sharing research data*. Natl Academy Press; Washington D.C.: 1985.
6. National Institutes of Health. [April 17, 2012] Final NIH Statement on Sharing Research Data. Notice NOT-OD-03-032. 2003; <http://grants.nih.gov/grants/guide/notice-files/not-od-03-032.html>.
7. Thomson Reuters. [October 29, 2012] Solving the Issues of Discovery, Attribution and Measurement in Data Sharing.. Collaborative Science Essay. 2012. http://wokinfo.com/products_tools/multidisciplinary/dci/collaborative_science_essay.pdf.
8. Tenopir C, Allard S, Douglass K, et al. Data sharing by scientists: practices and perceptions. *PLoS ONE*. 2011; 6(6):e21101. [PubMed: 21738610]
9. Hernan MA, Wilcox AJ. Epidemiology, data sharing, and the challenge of scientific replication. *Epidemiology*. Mar; 2009 20(2):167–168. [PubMed: 19234410]
10. Norman C. Sharing research data urged. *Science*. Aug 16.1985 229(4714):632. [PubMed: 17739366]
11. Ross JS, Lehman R, Gross CP. The importance of clinical trial data sharing: toward more open science. *Circ Cardiovasc Qual Outcomes*. Mar 1; 2012 5(2):238–240. [PubMed: 22438465]
12. Estabrooks, CA.; Romyn, DM. *Canadian Journal of Nursing Research*. Vol. 27. Spring; 1995. Data sharing in nursing research: advantages and challenges.; p. 77-88.
13. Hrynaskiewicz I, Altman DG. Towards agreement on best practice for publishing raw clinical trial data. *Trials*. 2009; 10:17. [PubMed: 19296844]
14. Insel TR, Volkow ND, Li TK, Battey JF Jr, Landis SC. Neuroscience networks: data-sharing in an information age. *PLoS Biology*. Oct.2003 1(1):E17. [PubMed: 14551914]
15. Tai B, Straus MM, Liu D, Sparenborg S, Jackson R, McCarty D. The first decade of the National Drug Abuse Treatment Clinical Trials Network: bridging the gap between research and practice to improve drug abuse treatment. *J Subst Abuse Treat*. Jun; 2010 38(Suppl 1):S4–13. [PubMed: 20307794]
16. Wu LT, Blazer DG, Woody GE, et al. Alcohol and drug dependence symptom items as brief screeners for substance use disorders: Results from the Clinical Trials Network. *J Psychiatr Res*. Mar; 2012 46(3):360–369. [PubMed: 22204775]
17. Wu LT, Swartz MS, Pan JJ, et al. Evaluating brief screeners to discriminate between drug use disorders in a sample of treatment-seeking adults. *Gen Hosp Psychiatry*. Jan; 2013 35(1):74–82. [PubMed: 22819723]
18. Wu LT, Ling W, Burchett B, et al. Use of item response theory and latent class analysis to link poly-substance use disorders with addiction severity, HIV risk, and quality of life among opioid-dependent patients in the Clinical Trials Network. *Drug and Alcohol Dependence*. Nov 1; 2011 118(2-3):186–193. [PubMed: 21501933]
19. Wu LT, Ling W, Burchett B, Blazer DG, Shostak J, Woody GE. Gender and racial/ethnic differences in addiction severity, HIV risk, and quality of life among adults in opioid

- detoxification: results from the National Drug Abuse Treatment Clinical Trials Network. *Substance Abuse and Rehabilitation*. 2010; (1):13–22. [PubMed: 21709734]
20. Pilowsky DJ, Wu LT, Burchett B, Blazer DG, Woody GE, Ling W. Co-occurring amphetamine use and associated medical and psychiatric comorbidity among opioid-dependent adults: results from the Clinical Trials Network. *Substance Abuse and Rehabilitation*. 2011; 2:133–144. [PubMed: 21886430]
 21. Wu LT, Pan JJ, Blazer DG, Tai B, Stitzer ML, Woody GE. Using a latent variable approach to inform gender and racial/ethnic differences in cocaine dependence: a National Drug Abuse Treatment Clinical Trials Network study. *Journal of Substance Abuse Treatment*. Jun; 2010 38(Suppl 1):S70–79. [PubMed: 20307798]
 22. Wu LT, Pan JJ, Blazer DG, et al. An item response theory modeling of alcohol and marijuana dependences: a National Drug Abuse Treatment Clinical Trials Network study. *J Stud Alcohol Drugs*. May; 2009 70(3):414–425. [PubMed: 19371493]
 23. Burlew AK, Weekes JC, Montgomery L, et al. Conducting research with racial/ethnic minorities: methodological lessons from the NIDA Clinical Trials Network. *Am J Drug Alcohol Abuse*. Sep; 2011 37(5):324–332. [PubMed: 21854274]
 24. Wu LT, Blazer DG, Patkar AA, Stitzer ML, Wakim PG, Brooner RK. Heterogeneity of stimulant dependence: a national drug abuse treatment clinical trials network study. *Am J Addict*. May-Jun; 2009 18(3):206–218. [PubMed: 19340639]
 25. Wu LT, Blazer DG, Stitzer ML, Patkar AA, Blaine JD. Infrequent illicit methadone use among stimulant-using patients in methadone maintenance treatment programs: a national drug abuse treatment clinical trials network study. *Am J Addict*. Jul-Aug;2008 17(4):304–311. [PubMed: 18612886]
 26. Wu LT, Pan JJ, Blazer DG, et al. The construct and measurement equivalence of cocaine and opioid dependences: a National Drug Abuse Treatment Clinical Trials Network (CTN) study. *Drug and Alcohol Dependence*. Aug 1; 2009 103(3):114–123. [PubMed: 19423244]
 27. Hartzler B, Donovan DM, Huang Z. Rates and influences of alcohol use disorder comorbidity among primary stimulant misusing treatment-seekers: meta-analytic findings across eight NIDA CTN trials. *Am J Drug Alcohol Abuse*. Sep; 2011 37(5):460–471. [PubMed: 21854291]
 28. Hartzler B, Donovan DM, Huang Z. Comparison of opiate-primary treatment seekers with and without alcohol use disorder. *Journal of Substance Abuse Treatment*. Sep; 2010 39(2):114–123. [PubMed: 20598831]
 29. Lindblad R, Campanella M, Styers D, Kothari P, Sparenborg S, Rosa C. Strategies for safety reporting in substance abuse trials. *Am J Drug Alcohol Abuse*. Sep; 2011 37(5):440–445. [PubMed: 21854288]
 30. Brooks A, Meade CS, Potter JS, Likhnygina Y, Calsyn DA, Greenfield SF. Gender differences in the rates and correlates of HIV risk behaviors among drug abusers. *Substance Use and Misuse*. Dec; 2010 45(14):2444–2469. [PubMed: 20536356]
 31. Korte JE, Rosa C, Wakim P, Perl H. Addiction treatment trials: how gender, race/ethnicity, and age relate to ongoing participation and retention in clinical trials. *Substance Abuse and Rehabilitation*. 2011; (2):205–218. [PubMed: 24474858]
 32. Greenfield SF, Rosa C, Putnins SI, et al. Gender research in the National Institute on Drug Abuse National Treatment Clinical Trials Network: a summary of findings. *Am J Drug Alcohol Abuse*. 2011; 37(5):301–312. [PubMed: 21854272]
 33. Korte JE, Magruder KM, Chiuzaan CC, et al. Assessing drug use during follow-up: direct comparison of candidate outcome definitions in pooled analyses of addiction treatment studies. *Am J Drug Alcohol Abuse*. 2011; 37(5):358–366. [PubMed: 21854278]
 34. Rosa C, Ghitza U, Tai B. Selection and utilization of assessment instruments in substance abuse treatment trials: the National Drug Abuse Treatment Clinical Trials Network experience. *Subst Abuse Rehabil*. Jul 17; 2012 3(1):81–89. [PubMed: 22904649]
 35. National Heart Lung and Blood Institute. [July 12, 2012] Requesting NHLBI Data Repository Data Sets through BioLINCC.. BioLINCC Frequently Asked Questions. <https://biolincc.nhlbi.nih.gov/faqs/#toc6>.

36. Center for Medicare & Medicaid Services. [July 11, 2012] Requesting CMS's Limited Data Set (LDS) Files.. Limited Data Set (LDS) Files [government document]. 2011. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/LimitedDataSets/Downloads/RevisedLDSInstructions.pdf>.
37. Center for Drug Evaluation and Research. [April 17, 2012] CDER Common Data Standards Issues Document (Version 1.1/December 2011). [Food and Drug Administration website]. 2011. <http://www.fda.gov/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/ucm248635.htm>.
38. National Institute on Drug Abuse (NIDA). [April 17, 2012] Notice Announcing Data Harmonization for Substance Abuse and Addiction via the PhenX Toolkit. [NIH website]. 2012. <http://grants.nih.gov/grants/guide/notice-files/NOT-DA-12-008.html>.

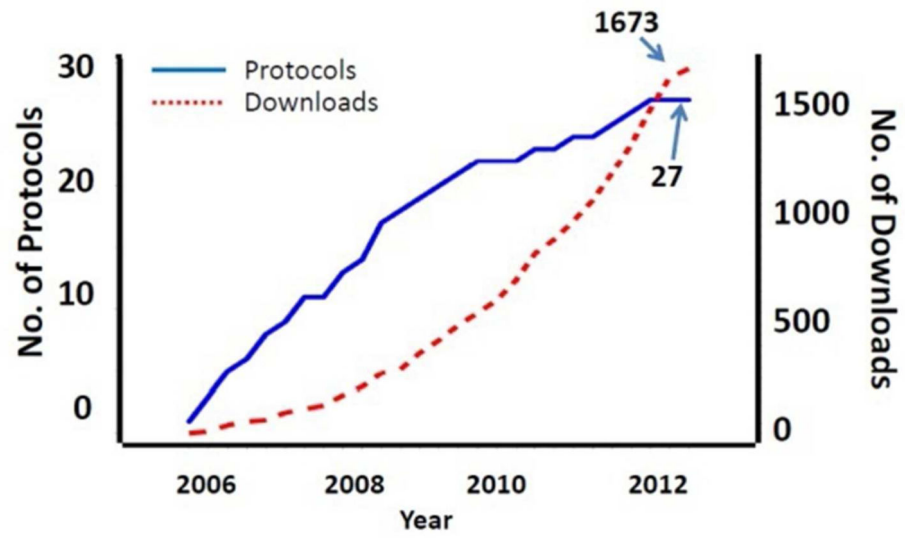


Figure 1.
Data Share website usage over time.

Table 1

Completed CTN Studies with Clinical Data Posted on the Data Share Website through July 2012

Study Number	Keywords	Number of Assessments	Date Posted to Website	Number of Downloaded Datasets by Public Users
NIDA-CTN-0001	Clonidine, Buprenorphine, Opiate, Naloxone	7	05/08/2006	255
NIDA-CTN-0002	Naloxone, Clonidine, Buprenorphine, Opiate	7	07/26/2006	89
NIDA-CTN-0003	Opiate, Naloxone, Buprenorphine, Suboxone	6	05/29/2008	69
NIDA-CTN-0004	Motivational Interviewing	9	10/03/2007	86
NIDA-CTN-0005	Motivational Interviewing	6	09/08/2006	74
NIDA-CTN-0006	Motivational Incentives	5	10/31/2006	66
NIDA-CTN-0007	Methadone, Motivational Incentives	4	12/13/2006	44
NIDA-CTN-0008	Community Treatment Programs	4	01/25/2007	45
NIDA-CTN-0009	Nicotine Replacement, Smoking	17	12/03/2007	72
NIDA-CTN-0010	Heroin, Buprenorphine, Adolescent, Naloxone	10	07/30/2009	57
NIDA-CTN-0011		2	05/18/2007	29
NIDA-CTN-0012	HIV/AIDS, Hepatitis C, Sexually Transmitted Infections	3	06/11/2007	34
NIDA-CTN-0013	Women, Motivational Interviewing	9	05/27/2008	42
NIDA-CTN-0014	Adolescent, Family Therapy	14	03/02/2010	30
NIDA-CTN-0015	PTSD, Women	14	05/01/2009	70
NIDA-CTN-0016		6	08/01/2007	15
NIDA-CTN-0017	Hepatitis C, HIV/AIDS	8	12/23/2008	34
NIDA-CTN-0018	HIV/AIDS, Men, Sexually Transmitted Infections	9	10/02/2008	40
NIDA-CTN-0019	HIV/AIDS, Sexually Transmitted Infections, Women	17	10/28/2008	64
NIDA-CTN-0020	Job Training	12	02/27/2009	35
NIDA-CTN-0021	Motivational Enhancement	12	07/30/2008	53
NIDA-CTN-0027	Opiate, Methadone, Buprenorphine	6	06/06/2012	3
NIDA-CTN-0028	ADHD, Adolescent, OROS-MPH	12	10/04/2010	49
NIDA-CTN-0029	ADHD, OROS-MPH, Smoking	16	12/15/2009	30
NIDA-CTN-0030	Buprenorphine, Naloxone, Opiate	12	06/22/2011	73
NIDA-CTN-0031	stimulant	12	02/06/2012	20
NIDA-CTN-0032	HIV/AIDS	10	11/02/2011	24

Table 2

Assessment Categories and Subcategories

Category (number of measures)	Subcategories	Example of Assessments
Demographics and Employment (7)		Vocational Survey Pre-treatment
Health Perceptions and Quality of Life (4)		EuroQol EQ-5D
Clinic Related Surveys (12)		Treatment Program Administrator Survey
Interpersonal Relationships and Culture (10)		Family Environmental Scale
Physical and General Health (8)		SF-36 Health Status Questionnaire
Substance Use (40)	Tobacco, Alcohol, Drugs	Addiction Severity Index
Sexual Behavior and HIV (18)		Attitudes Toward Condom Use
Impulsivity and General Trait and Behavior Scales (7)		Peer Delinquency Scale
Mental Health (22)	ADHD, Antisocial Personality, Anxiety, Dementia, Depression, Diagnostics, Eating Disorders, General/Multiple Disorders, PTSD, Suicidal Intent	Beck Depression Inventory
Total: 9 categories	13 subcategories	127 assessments