CONSIDER Statement: Consolidated Recommendations for Sharing Individual Participant Data from Human Clinical Studies

Craig S. Mayer, MS[1], Nick Williams, Ph.D[1], Vojtech Huser MD, Ph.D[1]
[1]Lister Hill National Center for Biomedical Communication, National Library of Medicine, NIH Bethesda, MD

Abstract

Many research sponsors require sharing of data from human clinical trials. We created the CONSIDER statement, a set of recommendations to improve data sharing practices and increase the availability and re-usability of individual participant data from clinical trials. We developed the recommendations by reviewing shared individual participant data and study artifacts from a set of completed studies, as well as study data deposited on ClinicalTrials.gov and on several data sharing platforms. The CONSIDER statement is comprised of seven sections including: format, data sharing, study design, case report forms, data dictionary, data de-identification and choice of data sharing platform. We developed several different forms of CONSIDER which includes a brief form (the checklist), a full form (detailed descriptions and examples), and a scoring methodology. The checklist can be used to evaluate adherence to various progressive data sharing recommendations. We are currently in Phase 2 of collecting feedback on the CONSIDER statement.

Introduction

In the past, influential policy documents that guide publishing of medical journal articles achieved advances on how results of clinical research are reported. For example, the CONSORT statement and checklist targeted improved reporting of participant flow diagrams and interventional trial designs, analysis and interpretations.[1] Similarly, the STROBE statement aimed to improve the reporting of observational studies.[2] The creation and use of these policy documents have had a profound positive effect in their respective areas.[3,4] We assume that same mechanism may help improve sharing of de-identified individual participant data (IPD) from completed human clinical studies (interventional trials and observational studies).

By policy, many research sponsors currently require sharing of data from human clinical studies.[5] In some of these cases there is not a consensus on a method or format for sharing IPD data. This results in Principle Investigators (PIs) and study sponsors having to decide how best to ensure they comply with this requirement.

For secondary data users there are many challenges when dealing with acquiring and using shared data. These challenges include data integration and interoperability issues, ethical and policy considerations, financial and time constraints and data quality and annotation problems.[6,7] The burden of many of these challenges are shared with PIs who are looking to make their data available, but PIs can mitigate these challenges by taking pre-emptive action when anticipating sharing their study data.[8] Despite the challenges, there are many benefits to sharing study data and data reuse, such as significant reductions in the time and funding that would otherwise be needed to produce new data, as well as the ability to generate and test new hypotheses and compare and contrast multiple sources of clinical trial data.[9,10]

There are many considerations PIs need to account for in order to minimize these challenges of sharing their data and maximize the capabilities of its re-use. These considerations include: what data and data artifacts should be shared, what formats study resources should be shared in, and how to best make accessible and share these resources, either via a data sharing platform or by request from interested secondary researchers.

There are different formats to be considered in study design when it comes to data collection and reporting. While in the past, many studies have used custom formats developed specifically for the study for data collection and reporting, there are various initiatives that have been developed over recent years that maximize the interoperability and reuse of collected participant data by being implemented during study design. These initiatives include the use of data standards or common data elements (CDEs) that aid in the harmonization of different clinical studies.[11] These efforts are to encourage and improve the re-use capabilities of clinical study data. The most popular of these initiatives are standards set by the Clinical Data Interchange Standards Consortium (CDISC), which despite existing for over two

decades, has not become widely adopted by academic medical centers and are mainly used by pharmaceutical research sponsors, thanks to the Food and Drug Administration (FDA) mandate.

As far as making data available for re-use there are many methods that can be chosen. Over the past few years several data sharing platforms with various capabilities have been developed to share and acquire IPD from completed human clinical trials. Some platforms such as NIDA Data Share, are domain specific, while others, such as Vivli or Clinical Study Data Request (CSDR), are general and include studies from a variety of clinical domains.

We present a set of recommendations to improve the practice of data sharing and promote data re-use, known as the Consolidated Recommendations for Sharing Individual Participant Data from Human Clinical Studies (CONSIDER statement). The acronym is loosely based on letters contained in the title: CONSolidated REcommendations for sharing Individual participant Data. Letters E and R are re-ordered to create a more memorable acronym. This set of recommendations provides a checklist and a set of recommendations to guide PIs, study team members, study sponsors and data sharing platform representatives for optimal clinical research data sharing. We use the term study to refer to both interventional trials and observational studies.

Methods

Set of reviewed studies

To develop the CONSIDER statement, we looked at the current state of sharing data from human clinical studies. Our analysis of shared study materials included reviewing the structure and features of shared IPD data and analyzing the presence and format of different study artifacts (e.g. data dictionary, case report forms [CRF], etc.). We also analyzed the accessibility and availability of study resources by reviewing and following the process for requesting study data on data sharing platforms and analyzing study records on a clinical trial registry where we identified which fields are commonly included and excluded by record administrators.

Our analysis included a set of HIV clinical studies which were obtained as part of a larger project focusing on CDEs in HIV studies.[12] While the trials underpinning the recommendations are primarily made up of HIV-related studies, the recommendations are not limited to HIV, and were generated as general recommendations that can be applied to studies from any clinical field. Further adding to the generalization of our recommendations, the data platforms and sources, as well as certain acquired data artifacts used in developing the recommendations are not HIV specific and include non-HIV studies and data sources that exemplify good data sharing practices or present popular challenges associated with data sharing.

The CONSIDER statement was developed from the review of IPD from 30 studies, study data artifacts from 48 studies, and an analysis of 10 data sharing platforms and 6 clinical trial networks. We also did a comprehensive review of clinical study registration data, the presence of data artifacts and the plan to share IPD for HIV trials on ClinicalTrials.gov.[13]

Design assumptions

The development and use of CONSIDER requires a few comments and assumptions. First, we do not recommend any one data sharing platform or data standard or structure. These recommendations are intended to recommend specific features and capabilities rather than a specific entity.

Second, the CONSIDER statement was also developed with the knowledge of the constraints and limited resources PIs face when it comes to staff, time, funding, and privacy. With that in mind, we know this limits the capabilities for a PI to prepare data for sharing, use a specific sharing platform or use certain data structures. The CONSIDER statement is intended to be used as a checklist to use the optimal practices, where plausible, to maximize the visibility, shareability and re-usability of clinical study data.

Results

Sections

Based on our analysis we structured CONSIDER into seven sections that reflect key areas relevant to data sharing. Each section contains between 1 and 13 checklist items. The sections are:

1. *Data Format:* Recommendations on the data structure and the inclusion of certain aspects and elements of the IPD. Using certain methods when formatting IPD can greatly improve the functionality of the data to data re-users and ensure an effective analysis of the shared data.

2. *Data Sharing:* Includes how to make the study available and visible to potential secondary researchers. This section also includes how to share information about the study that can give data re-users a complete understanding of the study and how best to use the available data.

3. *Study Design:* Includes data collection and data sharing recommendation that are important to consider by PIs during study design. The recommendations aim to improve the data usability and data comparability (to other similar studies).

4. *Case Report Forms:* Recommendations about the inclusion of CRFs. CRFs are valuable study artifacts for data recipients to understand the data collection process and the underlying documents that generated the data being analyzed.

5. *Data Dictionary:* Provides recommendations about the availability, format, and features to include when sharing the data dictionary of a study. This section includes making data dictionaries as widely and publicly available as possible and in a format that is easy to use (machine readable). This section also includes recommendations on key information about each data element or form to include in the dictionary. A more comprehensive analysis of data dictionaries leading to the recommendations included in this section can be seen in our previous work.[14]

6. *Data De-identification:* Focuses on describing how data was redacted during de-identification and what should be provided to data recipients when referring to de-identification techniques so they have an understanding of how the data has been changed from the raw collected data. Data de-identification also include the clear communication about the rights and restrictions of data recipients when referring to the process and risks of potential re-identification.

7. *Choice of a Data Sharing Platform:* Provides features and capabilities to look for when choosing a data sharing platform to deposit data (at study completion). The recommendations specify desired platform features such as the ability to quickly find relevant studies (search capabilities), the presence of available study resources and metadata, and the ability to quickly request and acquire available IPD. Different features and capabilities of data sharing platforms improve the effectiveness and efficiency for sharing IPD from clinical studies.

Recommendations

Table 1 shows CONSIDER as a list of recommendations (or checklist format) structured by the previously mentioned sections. For the full version of the CONSIDER statement go to w3id.org/CONSIDER.

Table 1. Individual CONSIDER recommendations by section.

Section	Title of recommendation
Format	Share person table in CDISC or OMOP format
	Group data and data elements into relevant data domains (e.g., medication history, laboratory results history, medical procedure history)
	Follow a convention when using relative time.
	Utilize previously defined Common Data Elements and reference them by their identifiers
	Use formats that can be natively loaded (without highly specialized add-ons) into multiple statistical platforms
Data Sharing	Register your study at ClinicalTrials.gov registry
	Do not limit study metadata to the legally required elements. Also populate optional elements (such as data sharing metadata)
	Fully populate data_sharing_plan text filed on ClinicalTrials.gov (if sharing data)
	If Individual Participant Data is shared on a data sharing platform, update the ClinicalTrials.gov record with the URL link to the data.
	Provide basic summary results using results registry component of Clinicaltrials.gov
	Utilize ClinicalTrials.gov fields for uploading study protocol, empty case report forms, statistical analysis plan and study URL link
	Provide de-identified Individual Participant Data
Study Design	Adopt previously defined applicable Common Data Elements
Case Report Forms	Share all Case Report Forms used in a study
	List all CRFs
Data Dictionary	Provide data dictionary
	Provide data dictionary in machine readable format
	Separate data dictionary from de-identified individual participant data. Since it contains no participant level data, do not require local ethical approval as a condition of releasing the data dictionary (avoid a requestwall for data dictionary).
	Share a data dictionary as soon as possible. Do not wait until the data collection is complete.
	Provide data dictionary in a single, machine-readable file.
	For each data element, provide a data type (such as numeric, date, string, categorical)
	For categorical data elements, provide a list of permissible values and distinguish when numerical code or string code is a code for a permissible value (versus actual number or string)
	Distinguish categorical string data elements from free-text string data elements
	Link utilized Common Data Elements adopted by your study to appropriate terminologies
	Link data elements or permissible values to applicable routine healthcare terminologies (either because you designed them to be linked or post-hoc, they can be semantically linked as equivalent)
	Provide complete data dictionary (all elements in data are listed in a dictionary) and all types of applicable dictionaries (date elements, forms [or groupings], and permissible values)
	Include sufficient description for data elements
	Use identifiers (unique where applicable) for data element, forms and permissible values.
Data de-identification	Provide data de-identification notes
Choice of a Data Sharing platform	Use platforms that allows download of all studies available on the platform
	Choose a platform that supports batch request (ability to request multiple studies with one request)

CONSIDER formats

We developed three views of the CONSIDER statement which have varying levels of detail and serve different purposes. The first format is the *brief view*, which is just a list of the different recommendations and their associated sections as seen in Table 1. The second format is the *full view* that on top of what is included in the brief format, also includes a detailed description of the recommendation, a positive example that features a study or platform that demonstrates full or partial compliance with the recommendation, and optionally a challenging example where the recommendation was not followed and what challenge that leads to during data re-use. For some recommendations, we used positive and challenging examples outside the input set of studies described in Methods. Finally, the third view is the *score sheet view* that assumes familiarity with the individual recommendations and is meant to facilitate individual study scoring. It lists each recommendation (by section) and the scoring instructions.

We acknowledge that some CONSIDER items depend on each other and can be considered partially overlapping. For example, the requirement to list data type for each data element in the data dictionary partially overlaps with properly handling categorical data elements. The description field (in the full view) for each CONSIDER item contains an explanation and rationale for this overlap. We chose to allow some overlap because we saw studies that formally comply with some recommendation, but closer scrutiny reveals additional deficiencies. The seemingly overlapping recommendations are meant to fully clarify and describe the best data sharing practices.

CONSIDER score and scoring approach

We developed a scoring system that scores each recommendation separately and then counts the score by each section. The higher the score the better the practices are implemented by the given study.

Each checklist item is scored on a zero to one scale. For binary items, the possible values are either one for practicing the recommendation or zero if it does not. For items where partial assessment of compliance is possible, a range of values between zero and one can be assigned (with 2-digit precision) depending on how completely the recommendation is followed. For example, for the recommendation 'for each data element provide a data type' if the study provides a data type for 47.15% of the data elements then it will get a score of 0.47 for that checklist item.

Certain recommendations included in CONSIDER rely on publicly available information from ClinicalTrials.gov (CTG) registry. To facilitate the easy application of the CONSIDER checklist, we created an R script that uses the relational database version of CTG, known as the Aggregated Analysis of ClinicalTrials.gov (AACT database; published and maintained by Duke University), to automatically score a subset of checklist items that can be assessed by CTG study registration metadata.[15] This script (located at w3id.org/CONSIDER) takes as input a set of study CTG identifiers (called NCTs) and returns their CONSIDER scores for the subset of checklist items that can be automated.

Example application of the CONSIDER checklist to individual studies

We applied CONSIDER to two trials from our set of analyzed trials to show how the checklist can be applied to individual studies. We scored NCT01751646 'Vitamin D Absorption in HIV Infected Young Adults Being Treated With Tenofovir Containing cART', which has IPD data deposited on the National Institute of Child Health and Human Development's Data and Specimen Hub platform (NICHD DASH), and NCT01233531 'Effects of Cash Transfer for the Prevention of HIV in Young South African Women', which has IPD available upon request from the HIV Prevention Trial Network (HPTN). Both are also registered at CTG. Table 2 shows the scores and percentages for these two studies when applying the CONSIDER checklist.

Table 2. Results of applying CONSIDER scoring for interventional trials NCT01751646 and NCT01233531.

Section	Best Possible Score	NCT01751646	NCT01233531
Format	5	3 (60.0%)	3 (60.0%)
Data Sharing	7	4 (57.1%)	3 (42.9%)
Study Design	1	0 (0.0%)	0 (0.0%)
Case Report Forms	2	2 (100.0%)	1 (50.0%)
Data Dictionary	13	9.45 (72.7%)	6.72 (51.7%)
Data de-identification	1	1 (100.0%)	0 (0.0%)
Choice of a Data Sharing platform	2	0 (0.0%)	2 (100.0%)

Discussion

Seeking feedback

We expect evolution of the CONSIDER checklist and we welcome feedback to any checklist item or section at craig.mayer2@nih.gov. CONSIDER was first developed in May 2019. We performed a Phase 1 feedback stage from Sept. 2019-Dec. 2019, where we elicited feedback from an internally selected group of experts. We are currently (since Jan. 2020) in Phase 2 of collecting feedback from the larger CRI community. As part of the Phase 2 feedback process, we created a mechanism and set of questions specific to the different perspectives involved in the data sharing process. This includes targeted feedback questions intended for PI's involved in study design, study data custodians involved in data housing and distribution, data sharing platform administrators, and data recipients. The developed targeted feedback tools are intended to better assist in understanding the capabilities and challenges for each individual involved in the data sharing process and allow us to better formulate well-rounded recommendations that are both feasible and beneficial for all involved.

Limitations

The resulting CONSIDER statement has several limitations. First, we focused on a US context and considered ClinicalTrials.gov registry. Second, we only used a limited set of studies to arrive at the recommendations. Using a larger set may result in a more comprehensive coverage of best practices. Third, the scoring system assumes equal importance (and weight) of each item. It would be feasible to develop a weighted score if agreement on prioritization can be reached. We also currently do not attempt to combine the score of individual CONSIDER sections into a single score. Fourth, we received only limited feedback from some stakeholder groups (platform administrators and PIs) and plan a focused feedback seeking campaign to address this.

Future work

To further develop CONSIDER we will continue to assess the state of data sharing and accept feedback to add recommendations and sections as they become necessary. We will also look to improve the capabilities of scoring individual studies by automating the process (as we have done with CTG related recommendations), which may include linking directly to other clinical trial registries, data sharing platforms and individual study pages. We also understand the different recommendations may present certain challenges and we intend to assess how demanding and resource intensive the implementation of each recommendation is to better recommend the most practical and implementable practices.

Conclusion

We analyzed data sharing platforms, data artifacts, study registration metadata, and shared IPD for completed clinical trials and created a set of recommendations, called the CONSIDER statement. The CONSIDER statement consists of seven key sections devoted to data format, data sharing, study design, CRFs, data dictionary, data de-identification, and choice of data sharing platform. These recommendations can be used to score existing studies to evaluate adherence to good data sharing practices. The recommendations can also be used to guide PIs and study sponsors to improve data sharing of future studies. We expect evolution of the CONSIDER statement based on input from clinical research informatics experts and the wider research community.

References

1. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. PLoS Med. 2010 Mar 24;7(3):e1000251.

2. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. Ann Intern Med. 2007 Oct 16;147(8):573.

3. Bastuji-Garin S, Sbidian E, Gaudy-Marqueste C, Ferrat E, Roujeau J-C, Richard M-A, et al. Impact of STROBE Statement Publication on Quality of Observational Study Reporting: Interrupted Time Series versus Before-After Analysis. Schooling CM, editor. PLoS ONE. 2013 Aug 26;8(8):e64733.

4. Poorolajal J, Cheraghi Z, Irani AD, Rezaeian S. Quality of Cohort Studies Reporting Post the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement. Epidemiol Health. 2011 Jun 7;33:e2011005.

5. Taichman DB, Sahni P, Pinborg A, Peiperl L, Laine C, James A, et al. Data sharing statements for clinical trials. BMJ. 2017 Jun 5;357:j2372.

6. Meystre SM, Lovis C, Bürkle T, Tognola G, Budrionis A, Lehmann CU. Clinical Data Reuse or Secondary Use: Current Status and Potential Future Progress. Yearb Med Inform. 2017;26(01):38–52.

7. Wilkinson T, Sinha S, Peek N, Geifman N. Clinical trial data reuse – overcoming complexities in trial design and data sharing. Trials. 2019 Dec;20(1):513.

8. Mbuagbaw L, Foster G, Cheng J, Thabane L. Challenges to complete and useful data sharing. Trials. 2017 Dec;18(1):71.

9. Kawahara T, Fukuda M, Oba K, Sakamoto J, Buyse M. Meta-analysis of randomized clinical trials in the era of individual patient data sharing. Int J Clin Oncol. 2018 Jun;23(3):403–9.

10. Rosenblatt M, Jain SH, Cahill M. Sharing of Clinical Trial Data: Benefits, Risks, and Uniform Principles. Ann Intern Med. 2015 Feb 17;162(4):306.

11. Sheehan J, Hirschfeld S, Foster E, Ghitza U, Goetz K, Karpinski J, et al. Improving the value of clinical research through the use of Common Data Elements. Clin Trials J Soc Clin Trials. 2016 Dec;13(6):671–6.

12. Huser V, Mayer CS, Williams N. Real World Data and Research Common Data Elements: a Case Study in HIV. AMIA Clinical Informatics Conference, May 2020.

13. Huser V. Sharing of de-identified patient level data from human clinical trials: analysis of US-based studies in the ClinicalTrials.gov registry. In 2017.

14. Mayer CS, Williams N, Huser V. Analysis of data dictionary formats of HIV clinical trials. Huy NT, editor. PLOS ONE. 2020 Oct 5;15(10):e0240047.

15. AACT Team. Aggregate Analysis of ClinicalTrials.gov (AACT) database [Internet]. 2020 [cited 2020 Mar 20]. Available from: https://aact.ctti-clinicaltrials.org/learn_more