

Sharing of Individual Participant Data from Clinical Trials: General Comparison and HIV Use Case

Craig S. Mayer, MS^{1,2}, Nick Williams, Ph.D^{1,2}, Sigfried Gold, MFA, MA¹,
Kin Wah Fung, MD, MS¹, Vojtech Huser MD, Ph.D¹

¹Lister Hill National Center for Biomedical Communication, National Library of Medicine, NIH
Bethesda, MD

²ICF Incorporated, Rockville, MD

Abstract

Sharing of individual participant data is encouraged by the International Committee of Medical Journal Editors. We analyzed clinical trial registry data from ClinicalTrials.gov (CTG) and determined the proportion of trials sharing de-identified Individual Participant Data (IPD). We looked at 3,138 medical conditions (as Medical Subject Heading terms). Overall, 10.8% of trials with first registration date after December 1, 2015 answered 'Yes' to plan to share de-identified IPD data. This sharing rate ranges between 0% (biliary tract neoplasms) to 72.2% (meningitis, meningococcal) when analyzed by disease that is focus of a study. Via a predictive model, we found that studies that deposited basic summary results data to CTG results registry, large studies and phase 3 interventional studies are most likely to declare intent to share IPD data. As part of an HIV common data element analysis project, we further compared a body of HIV trials (24% sharing rate) to other diseases.

Introduction

Sharing of de-identified individual participant data¹ is encouraged by the International Committee of Medical Journal Editors (ICMJE).² Since 2019, authors that publish trial results in an ICMJE member journal (more precisely for trials that begin enrolling participants after Jan 1, 2019) are required to include a clear data sharing statement in the trial's registration data². This does not mandate actual individual participant data (IPD) sharing; it only mandates describing whether IPD data will or will not be shared. However, many funders of clinical studies have a separate policy that promotes actual data sharing.^{3,4} Data sharing is closely related to a trial registration mandate required for applicable trials in the US. ClinicalTrials.gov (CTG) is the world's largest registry of interventional trials and observational studies⁵. CTG study record administrators are encouraged to keep the study record in the CTG registry current in order to facilitate public review of ongoing or completed studies. Since 2015, CTG allows study record administrators to specify an IPD data sharing plan.

We investigate the question of data sharing from the perspective of a study sponsor that is highly motivated to share data and consequently keeps the CTG study record current. Since 2017, CTG also added the ability to provide links to study protocol, Case Report Forms (CRFs), other study documents, or IPD. In case of documents, these can be uploaded directly to CTG or provided via a link. For IPD, only links are accepted. CTG's ability to refer to other trial results platforms made CTG a central place and a good starting point to not only discover relevant clinical studies but to obtain important study documentation and IPD. We assumed that motivated study record sponsors would take advantage of CTG serving as the starting point to provide study metadata to the public. We fully acknowledge that the ability to provide a data sharing plan and provide an external link to de-identified IPD was only added in December of 2015 and trials that did first registration prior to this date were unable to do so at the time. However to assess the impact of the timing issue, we looked at how often such trials update the IPD data sharing plan retrospectively (during a registry record update). As of December 20, 2018, we found that a total of 18,457 studies (6.3% of 292,680 registered studies at that time) that initially registered when specifying a data sharing plan was not possible (prior December 1, 2015), that retrospectively answered 'Yes', 'No' or 'Undecided' to the question about an IPD data sharing plan. A total of 4,134 studies answered 'Yes', 10,379 answered 'No' and 3,944 answered 'Undecided'. This finding shows that thousands of trials indeed update their CTG registration record, and it partially confirms its importance as a central public record.

In this study, we analyze study data sharing by disease domain. As part of a larger project focusing on analysis of clinical trials and their common data elements for the HIV and AIDS domain, we further focus on describing the body of all HIV trials (we use the term HIV trials to refer to both HIV and AIDS trials) and compare patterns of data sharing in HIV with other medical domains.

Methods

Determining data sharing on disease level

We used relational database representation of CTG study registration data called Aggregate Analysis of Clinical Trials.gov (AACT) created at Duke University.⁶ The AACT database is created by parsing XML representation of each study from CTG. We adopt the CTG terminology and use the term *study* to refer to both an *interventional trial* and an *observational study*. We used SQL language to query and extract data from AACT and R language to analyze the data. The study repository at <https://github.com/lhncbc/r-snippets-bmi/tree/master/CTG/sharing> contains selected analytical code and supplementary result files.

To assign a given trial to a disease, we used CTG's condition field. Each CTG study must specify a study condition that is defined as: 'The name(s) of the disease(s) or condition(s) studied in the clinical study, or the focus of the clinical study.'⁷ For example, the TOPKAT trial (NCT01352247) has 'Knee Osteoarthritis' as a single specified study condition. CTG aggregates similar condition entries into higher level groupings based on Medical Subject Headings (MeSH) terminology (CTG field: condition_mesh).⁸ The majority of trials (40.5%) have a single condition specified per study when all CTG studies are analyzed. A total of 29.9% of trials have two conditions and 29.6% have three or more conditions specified. Our intent was to analyze sharing by MeSH keyword, so studies that specified multiple keywords were included in any analysis for each specified keyword. To avoid bias in our analysis of overall sharing, we removed conditions that were considered outliers which had less than 15 registered studies.

To analyze data sharing intent, we used CTG's field called plan_to_share_IPD. During study registration, the study record administrator must indicate whether there is a plan to make individual participant data (IPD) collected in the study, including data dictionaries, available to other researchers (typically after the study is completed and a possible data embargo period has expired). The data sharing plan has 'Yes', 'No' and 'Undecided' as possible answers. To avoid bias, we only looked at studies which first registered after December 1, 2015 when this feature was available during first registration, studies that provided an answer in the plan_to_share_IPD field, and studies that had an enrollment greater than zero (we refer to this set of studies as Set_A). For analysis purposes we viewed 'Undecided' as not stating that the study plans to share and so included it as a 'No' answer. A related second field in CTG is plan_to_share_IPD_description, which is a free text field where details of the plan are described. If the answer was 'No' or 'Undecided', the description field can be used to explain why IPD will not be shared or why it has not been decided yet.

Predicting Sharing

To explore what factors influence a decision to share study IPD, we developed a predictive model. The model used logistic regression analysis due to the categorical nature of the outcome variable. The predicted outcome variable was the answer of 'Yes' to the plan_to_share_IPD field in CTG. In order to select predictors, we first conducted a descriptive analysis of available CTG study level metadata for studies that answered 'Yes'. Based on this analysis, we subsequently selected a set of predictors either used in past analyses of CTG data⁹ or identified in a review by two clinical informatics experts. The prediction model was done in an exploratory mode. A comprehensive model was out of scope.

HIV use case

Because of a larger project that focuses on common data elements in HIV, we analyzed as a case study the HIV body of trials. We did several HIV-specific analyses.

Sharing ratio: First, we created a set of MeSH terms that belong to the HIV/AIDS domain (reviewing only MeSH terms present in CTG) and compared HIV sharing ratios to other diseases.

Manual review of data sharing plan description: Second, we analyzed more closely the other data sharing CTG fields for HIV trials. CTG specifies the plan description should describe what specific individual participant data sets will be shared and how they will be shared. We only did this type of review for HIV studies (due to feasible sample size) and not for all CTG studies in Set_A. An example of a plan description is 'de-identified participant level data will be available through the clinicalstudydatarequest.com platform upon request'. We manually reviewed plan descriptions (CTG field name: plan_to_share_IPD_description) for the 69 HIV studies that said 'Yes' to plan_to_share_IPD with a goal to assess how well the meaning of the IPD sharing fields were being understood by CTG study record administrators and how often a well specified plan is being articulated. Elements of a good plan were clearly specified by ICMJE.² We also looked at instances and common answers where the fields were possibly misunderstood and

recommended elements were not clearly present. This included descriptions such as ‘sharing will be done via publications and conferences’. We classified each data sharing plan description (based on manual review) as either well specified or possibly misunderstood.

Predicting sharing (for HIV): We used the same predictive model approach to assess trial factors that predict IPD data sharing for HIV trials. We contrasted how HIV-specific model predictors are different from the general model (for all trials).

Results

We queried the AACT database on December 20, 2018 and analyzed 287,626 studies that were registered on CTG at that time. Of those registered studies, 74,582 provided an answer to the plan_to_share_IPD question. Overall, 9,435 studies (out of 287,626, 3.3%) are sharing IPD (without regard to first submission date or MeSH disease term for study focus), 46,017 studies (16.0%) are not sharing IPD and 19,130 studies (6.7%) were undecided on sharing IPD.

Determining data sharing on disease level

When we filtered the above set of studies with our selection criteria of (1) only include those registered after December 1, 2015 (the date when the CTG registry introduced additional data sharing questions), (2) that provided some answer (Yes/No/Undecided) to plan_to_share_IPD, and (3) had listed the enrollment of greater than zero, the number of analyzed studies reduced to 62,166 studies (21.6% out of 287,626; referred to as set_A). Supplemental file S1_diseases.xlsx (Tab1-all_keywords) at our study repository shows sharing ratio for all 3,138 MeSH keywords in this set of studies. The overall sharing ratio (without regard to MeSH term) for studies meeting these criteria is 10.8%. To better review sharing rate across diseases, we have removed outlier MeSH disease terms that have fewer than 15 total studies (e.g., adenocarcinoma in situ and coronary aneurysm; see Tab2-without-rare-MeSH-terms in the S1 file). The two MeSH terms with the highest sharing ratios were ‘meningitis, meningococcal’ (sharing rate of 72.2%) and dengue (61.1%). The lowest sharing ratio was zero, where no studies for those MeSH terms planned to share IPD. Zero sharing rate was observed for 70 MeSH terms. Example MeSH terms with a zero sharing rate and the highest total count of studies were biliary tract neoplasms (0 sharing out of 51 total studies), bone diseases (0 sharing out of 43 total studies), and obstetric labor, premature (0 sharing out of 39 total studies). A review of sharing rate and total count of studies for MeSH keywords with the largest count of total studies (after December 1, 2015), reveals the following MeSH keywords: diabetes mellitus (16.5% sharing out of 1,656 total studies), breast neoplasms (9.9% sharing out of 1,068 total studies) and depression (13.6% sharing out of 825 total studies).

Table 1. Share rate of a subset of diseases (by MeSH term)

MeSH Term	Count of studies sharing data	Total count of eligible studies	Sharing ratio (%)
meningitis, meningococcal	13	18	72.2 %
dengue	11	18	61.1 %
hemophilia b	11	24	45.8 %
carpal tunnel syndrome	21	64	32.8 %
eczema	40	169	23.7 %
diabetes mellitus	273	1656	16.5 %
arthritis, rheumatoid	43	286	15.0 %
depression	112	825	13.6%
asthma	33	327	10.1 %
breast neoplasms	106	1068	9.9%
biliary tract neoplasms	0	51	0.0 %
bone diseases	0	43	0.0 %
obstetric labor, premature	0	39	0.0 %
oropharyngeal neoplasms	0	35	0.0%

Predicting sharing

Descriptive analyses that were executed in support of selecting appropriate predictors are available in supplemental file S2-sharing-descriptive-results. The predictive model was executed for all studies in set_A with the predicted

outcome being an answer of ‘Yes’ to plan_to_share_IPD. The answer of ‘Undecided’ was essentially treated in the same group as an answer of ‘No’. Table 2 lists the relative effect of the predictors (their coefficients) and the predictor type (e.g., binary, categorical, continuous). Study sample size (CTG uses term enrollment to refer to study size in terms of number of enrolled participants) was converted into a categorical variable based on bin thresholds determined in the preparatory descriptive study (see section Study size in supplemental file S2). Large studies were considered having enrollment over 200, medium studies having enrollment of between 51 and 200 and small studies having enrollment of 50 or less.

To assess the predictors, the absolute value indicates its importance in the overall model. The positive sign of a predictor indicates higher probability of IPD sharing whereas negative sign indicates lower probability. Table 2 shows that the highest positive predictor is study size (studies with 201+ participants; coefficient 1.2) The second highest positive coefficient overall (has_results: Yes) was if the study has basic summary results deposited in CTG results registry. The highest negative predictors are sponsor types (categorized under Agency_class) of Other and U.S. Federal Agency. Outside of these sponsor types the two predictors with the largest negative coefficients, indicating a lower probability of sharing IPD were an observational study type and if the phase is listed as N/A, which includes studies on devices and behavioral interventions. While study type did not have as large an effect as other analyzed predictors on whether a study would share, observational studies and patient registries do have a negative effect on the probability of a study sharing.

In terms of terms of study sponsor, NIH sponsored studies are the most likely to share and has the highest positive effect of any sponsor type on the probability of a study sharing (0.296). The only other sponsor type with a positive effect on sharing is Industry (0.031). The other two sponsor types, U.S Federal agencies (non-NIH) and Other, had the greatest negative coefficients overall, indicating a significantly lower probability of IPD sharing by studies with these sponsors. Other, which includes foreign institutions, universities, hospitals and non-profit organizations, had the highest negative effect (-0.617) while U.S. Federal agencies, such as the VA and FDA had the second highest negative effect (-0.556). One surprise in the result or the coefficients of the predictor was that a study being phase 1 was not a negative predictor, but rather had the highest positive effect of any phase on the probability of a study sharing.

We executed the prediction model for each MeSH term and the results are available in supplemental table S3-prediction-by-MeSH-term.

Table 2. Prediction model results for all studies showing predictor coefficients (ordered by absolute value)

General Model Predictor	Predictor Coefficient	Predictor Type
Enrollment: 201+	1.207541	categorical(3 values)
Has_results: Yes	0.977774	binary
Agency_class: Other	-0.617322	categorical(4 values)
Agency_class: U.S. Fed	-0.556474	categorical(4 values)
Phase: N/A	-0.396123	categorical(7 values)
Study_type Observational	-0.348703	categorical(3 values)
Enrollment: 51-200	0.325061	categorical(3 values)
Phase: 1/2	-0.322666	categorical(7 values)
Agency_class: NIH	0.296163	categorical(4 values)
Phase: 1	0.248599	categorical(7 values)
Phase: 2/3	-0.245158	categorical(7 values)
Phase: 3	0.227539	categorical(7 values)
Study_type: Observational [Patient Registry]	-0.142474	categorical(3 values)
Phase: 4	-0.125912	categorical(7 values)
Phase: 2	-0.104619	categorical(7 values)
Start_date	-0.071769	continuous
Agency_class: Industry	0.031220	categorical(4 values)
Enrollment: 1-50	-0.011347	categorical(3 values)

HIV use case

Sharing ratio: Table 3 shows all HIV/AIDS MeSH terms found in CTG and their sharing ratios. It shows that for the largest HIV MeSH term (HIV infections) the sharing rate was 24.5% which is more than double the average rate of

10.8% observed across diseases. The sharing rate for HIV places it in the top 5 percentile when looking at sharing ratio of all MeSH terms (when outliers are removed).

Table 3. Share rate for all HIV MeSH terms

HIV MeSH Term	Count of studies sharing data	Total count of eligible studies	Sharing ratio (%)
AIDS-related opportunistic infections	11	24	45.8
HIV seropositivity	7	25	28.0
HIV infections	69	284	24.3
AIDS dementia complex	1	6	16.7
HIV-associated lipodystrophy syndrome	0	1	0.0
lymphoma, aids-related	0	1	0.0

Manual review of data sharing plan description: Since the inclusion of the plan to share IPD field in CTG there have been 69 studies with HIV related MeSH terms that have answered ‘Yes’ to plan_to_share_IPD for which we manually classified their data sharing plan descriptions. Of the 69 studies reviewed, 51 had well specified plans to share IPD, in which a specific platform for data sharing was stated or a process for which researchers can follow to obtain study data was articulated. There were 18 studies that had plans listed that were not well specified and were possibly misunderstood. Examples of common answers fitting this possibly misunderstood category were: (1) the field was left blank, (2) a plan will be developed in the future, (3) data will be shared via publications and/or presentations, and (4) data will be shared with the participants and investigators in the study. All of which are examples of plans that do not properly articulate a method to share IPD with interested parties and external researchers. In our 2017 prior work (across all diseases, not just HIV), we also noted that the plan was describing summary data sharing and not participant level data, which is another possible misunderstanding.¹⁰ The full list of the 18 misunderstood descriptions is available in the supplemental file S4-plans. CTG also allows trial administrators who answer ‘No’ or ‘Undecided’ to include an explanation of why they answered one of these two options, in the plan description field. Of the 215 HIV studies that answered ‘No’ or ‘Undecided’, 22 included an explanation in the plan_to_share_IPD_description field. Further showing possible misunderstanding of the IPD sharing fields, 7 of the 22 had descriptions of sharing plans that were well specified and included processes for outside researchers to obtain IPD. Such plan descriptions seem to imply ‘Yes’ to sharing; however, official CTG data had ‘No’ or ‘Undecided’ recorded.

Predicting sharing (for HIV): We executed the model using just the HIV set of studies as input and reviewed the results, focusing on the predicted effects of each variable for HIV studies compared to those of the overall model presented earlier.

Table 4. Prediction model results for HIV studies showing predictor coefficients (ordered by absolute value)

HIV Model Predictor	Predictor Coefficient	Predictor Type
Has_results: Yes	9.013530	binary
Enrollment: 201+	2.008813	categorical(3 values)
Enrollment: 51-200	1.398866	categorical(3 values)
Study_type: Observational [Patient Registry]	-0.999999	categorical(3 values)
Agency_class: NIH	-0.970701	categorical(4 values)
Agency_class: Other	-0.867272	categorical(4 values)
Phase:4	-0.686195	categorical(7 values)
Phase:2	-0.565848	categorical(7 values)
Study_type: Observational	-0.537078	categorical(3 values)
Phase: N/A	-0.497471	categorical(7 values)
Phase:1	-0.442918	categorical(7 values)
Phase:3	-0.422930	categorical(7 values)
Phase: 1/2	-0.352964	categorical(7 values)
Agency_class:U.S. Fed	-0.336885	categorical(4 values)
Start_date	0.196993	continuous
Phase: 2/3	-0.051399	categorical(7 values)
Agency_class: Industry	0.040073	categorical(4 values)
Enrollment: 1-50	-0.038142	categorical(3 values)

For HIV related studies, shown in Table 4, we found the most important predictors of sharing to be similar to those of the general model. The two highest positive coefficients were study size (studies with 201+ participants; coefficient of 2.01) and whether a study had basic summary results deposited in CTG. Compared with the general model, having summary results deposited (and several other predictors) had a much higher absolute value ('has_results: Yes' had 9.0 absolute value for HIV vs. 0.97 for the general model).

There were also some differences in the general versus HIV prediction model results. Certain factors that had a positive effect in the general model had a negative effect in the HIV model. This included the factors of both Phase 1 and Phase 3 as well as NIH sponsored studies. The largest negative factor on whether an HIV study would share IPD data was whether the study type was an observational patient registry. This is not a surprise as one major factor in sharing IPD is the completion of the study. Since patient registries do not have a limited temporal scope (are long term research projects) and are always actively adding data, the IPD dataset is technically never complete and has a different policy and sharing pattern than a typical interventional trial or time-limited observational study.

Another noticeable difference for the HIV model is the positive effect start date has on the sharing of HIV data compared to the minimal weight start date had on the general model, indicating a trend in increased sharing of HIV studies over time.

Discussion

Our study is the first to analyze intent to share IPD data by disease using CTG registration data. We show that there are significant differences in IPD sharing ratios across diseases (using MeSH terms). We also show different factors or predictors that impact IPD data sharing. The fact that studies that deposited basic summary results to CTG result registry are more likely to share IPD data is helpful in our larger project where we seek IPD data (or data dictionaries) for a large set of trials in a given disease (e.g., HIV,¹¹ asthma or lymphoma). For trials where no answer to the question about IPD data sharing on CTG was provided (but basic summary results were deposited), a data outreach effort that first targets large trials with basic summary results may be the most pragmatic approach. Our study also shows that HIV studies share IPD at a higher rate (24.3%) than a general set of studies from the same time period (10.8%). HIV studies have a higher sharing rate despite the fact that the HIV-infected population may be considered a sensitive population.

Relevant previous studies

Prior to our analysis, we considered previous relevant literature on CTG data analysis and data sharing. Chen et al. analyzed digital health studies in CTG.¹² Other analysis by Tse et al. and Durham look at challenges associated with CTG data and made recommendations on the use of CTG data. Stergiopoulos et al. reviewed the overall completeness of CTG records and the information present on CTG.^{9,13,14} Finally, Federer et al. analyzed data availability statements on a journal article level in a single journal PLOS ONE.¹⁵ To our knowledge, our study is the first to analyze (and compare across diseases) sharing of IPD data on study level using CTG registry data.

Limitations

The results presented have several limitations. First, in our analysis of sharing by disease, we used MeSH keywords and treated each MeSH term as a unique condition. We did not aggregate similar MeSH terms into possible higher level disease categories. This would require additional grouping knowledge base which we did not want to custom develop. This multiple keywords per disease can be observed, for example, on HIV keywords (see table 3) showing similar entries such as 'hiv infections' and 'hiv seropositivity'. For HIV use case analysis, we also kept results on MeSH term level (without aggregation) and considered the term with the most studies to be representative of the domain. We also observed that CTG assigns some studies MeSH terms (as disease focus of the study) that are too broad (e.g., 'syndrome' or 'disease'). We did not try to correct this misclassification. Second, when using the CTG field study_condition and assigning studies to a MeSH term, we did not try to select a single MeSH keyword for a study. We simply counted a study under each of its keyword. This double counting may be considered incorrect by some experts, however we chose to respect multiple keywords for a study as entered by trial record administrators. In other words, our chosen unit of analysis was MeSH keyword and not a study. Third, for our prediction model, we only used predictors available in CTG and picked predictors based on descriptive analysis, prior literature and expert opinion. A comprehensive prediction analysis could possibly incorporate a larger set of predictors.

Conclusion

We found that 10.8% of studies plan to share IPD data (as declared on CTG) and compared data sharing ratio by disease. The most important study features that predict future IPD data sharing are large sample size and deposition of basic summary results. As a case study, we analyzed more closely sharing rate and detailed plan description for HIV/AIDS studies.

Acknowledgement: This research was supported by the Intramural Research Program of the National Institutes of Health (NIH)/ National Library of Medicine (NLM)/ Lister Hill National Center for Biomedical Communications (LHNCBC) and NIH Office of AIDS Research. The findings and conclusions in this article are those of the authors and do not necessarily represent the official position of NLM, NIH, or the Department of Health and Human Services.

References

1. Ohmann C, Banzi R, Canham S, Battaglia S, Matei M, Ariyo C, et al. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. *BMJ Open*. 2017 Dec 14;7(12):e018647.
2. Taichman DB, Sahni P, Pinborg A, Peiperl L, Laine C, James A, et al. Data sharing statements for clinical trials. *BMJ*. 2017 05;357:j2372.
3. Policy on clinical trials | Wellcome [Internet]. [cited 2019 Mar 12]. Available from: <https://wellcome.ac.uk/node/1939>
4. NOT-OD-16-149: NIH Policy on the Dissemination of NIH-Funded Clinical Trial Information [Internet]. [cited 2019 Mar 12]. Available from: <https://grants.nih.gov/grants/guide/notice-files/not-od-16-149.html>
5. Huser V, Cimino JJ. Evaluating adherence to the International Committee of Medical Journal Editors' policy of mandatory, timely clinical trial registration. *J Am Med Inform Assoc JAMIA*. 2013 Jun;20(e1):e169-174.
6. AACT database download [Internet]. Available from: <https://aact.ctti-clinicaltrials.org/download>
7. ClinicalTrials.gov. Definitions [Internet]. Available from: <https://prsinfo.clinicaltrials.gov/definitions.html>
8. Medical Subject Headings [Internet]. Available from: <https://www.nlm.nih.gov/mesh>
9. Durham TA. How Did These Data Get Here? Recommendations for the Analysis of Data From ClinicalTrials.gov. *Ther Innov Regul Sci*. 2018 Nov 9;2168479018811825.
10. 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.

11. Website for research project: Identification of Research Common Data Elements in HIV/AIDS using data science methods [Internet]. Available from: <https://github.com/lhncbc/CDE/tree/master/hiv/>
12. Chen CE, Harrington RA, Desai SA, Mahaffey KW, Turakhia MP. Characteristics of Digital Health Studies Registered in ClinicalTrials.gov. *Characteristics of Digital Health Studies Registered in ClinicalTrials.gov Letters*. 2019 Feb 25 [cited 2019 Mar 7]; Available from: <https://dx.doi.org/10.1001/jamainternmed.2018.7235>
13. Tse T, Fain KM, Zarin DA. How to avoid common problems when using ClinicalTrials.gov in research: 10 issues to consider. *BMJ*. 2018 May 25;361:k1452.
14. Stergiopoulos S, Getz KA, Blazynski C. Evaluating the Completeness of ClinicalTrials.gov. *Ther Innov Regul Sci*. 2018 Jul 26;2168479018782885.
15. Federer LM, Belter CW, Joubert DJ, Livinski A, Lu Y-L, Snyders LN, et al. Data sharing in PLOS ONE: An analysis of Data Availability Statements. *PLOS ONE*. 2018 May 2;13(5):e0194768.