Journal of Cancer Clinical Trials

Tsai, J Cancer Clin Trials 2016, 1:3

Commentary Open Access

Sharing Clinical Trial Data

Kao-Tai Tsai*

Frontier Informatics Service, New Jersey, USA

***Corresponding author:** Kao-Tai Tsai, Frontier Informatics Service, New Jersey, USA, Tel: 9084647186; E-mail: tsai0123@yahoo.com

Received date: July 08, 2015; **Accepted date:** July 22, 2016; **Published date:** July 29, 2016

Why Sharing Clinical Trial Data

Discovery of new medicines to maintain the overall health of the general public has always been among one of the important goals in human history. With advanced technologies in all scientific arenas currently, this activity has been elevated to an unprecedented level with many newly discovered drugs to treat diseases which were previously thought to be untreatable.

Large number of institutions founded by government, academia, and private industry are engaging in drug development research ranging from basic science, drug safety studies on animals and human, to clinical trials on patients. As a consequence, huge amounts of data have been generated, which are often beyond the limited capacity of these institutions to perform in-depth data analyses or data mining to discover the hidden information in this voluminous repository of data.

For example, in the pharmaceutical industry, after analyzing several pre-specified primary and secondary endpoints for the purpose of submitting marketing applications of new drugs to the health authorities or for journal publications, the large amount of data generated, including both drug efficacy and safety, often are not analyzed again even though they were collected at great expense in human and financial resources. Moreover, when the number of patients with a certain disease is limited and multiple companies are conducting similar trials with same patient population, the number of patients in each study can be quite small. If data from these small to moderate size studies were to be shared, one would have a better opportunity to understand the disease more comprehensively, thereby increasing the opportunities to design better treatments for these patients.

One way to better utilize the data in these scenarios mentioned above is through data sharing so that additional researchers around the world can contribute their expertise to analyze and advance the drug discovery process. Data sharing can make data from basic science research and clinical trials available to other investigators for secondary research, including carrying out additional analyses, analyzing unpublished data, reproducing published findings, and conducting exploratory analyses to generate new research hypotheses.

Responsible data sharing, including both clinical and non-clinical is indeed at a very high level of public interest. It maximizes the contributions made by trial participants to scientific knowledge that benefit both patients and society as a whole. However, to design a responsible data sharing process to benefit drug discovery research and also to protect the confidential commercial information of the sponsors and privacy of trial participants is a formidable undertaking. The challenge is in setting clear practical expectations for data sharing, and with consensus on approaches in a responsible manner to mitigate the risks involved.

Stakeholders have concerns about data sharing which often focus on insufficient protection of their proprietary commercial or intellectual property information. Similarly, trial participants seek assurances that their data will be shared in a way that protects their privacy under the study informed consent agreement.

Since an important goal of responsible data sharing is to increase scientific knowledge that leads to better therapies for patients, the fundamental principles of data sharing need to include, among other things, maximizing the benefits of clinical trials while minimizing the risks of sharing data, respecting the privacy of participants whose data are to be shared, increasing public trust in clinical trials and the sharing of trial data, conducting data sharing in a fair manner, and striving for the improvement of the health of general public through research based on shared data.

Therefore, as recommended by the Institute of Medicine [1], and others, that trial sponsors should foster a culture in which data sharing is the expected norm, and should commit to responsible strategies aimed at maximizing the benefits, minimizing the risks, and overcoming the challenges of sharing data for all parties.

Access to Clinical Trial Data

To implement clinical data sharing and make it practical, arrangements for determining access to trial data need to balance several goals including protecting the privacy of research participants, reducing the likelihood of invalid analyses or misuse of shared data, avoiding undue burdens on seeking data access for secondary research, avoiding undue inconvenience to investigators and sponsors that share data, and enhancing public trust in data sharing.

In protecting the privacy of research participants, several methods to de-identify personal information are available [2-4] although they all have limitations. In addition, different jurisdictions may have different de-identification requirements. Moreover, the risk of re-identification depends on the context in which data are released, the type of data, and any additional information from other sources, for example, from social networks or some public/semi-public databases, that could be combined with shared data to re-identify a participant. This kind of data linkage can substantially increase the risk of re-identification of trial participants. In the case of genome-wide sequencing data analyses, for example, de-identification and data security alone may not provide adequate protection.

To further reduce the risk of re-identification, in addition to de-identifying data, other mechanisms are crucial to minimize these risks and related disincentives for data sharing. Such mechanisms may include employing data use agreements that include provisions to protect clinical trial participants, providing credit to investigators who collect the trial data, protecting the intellectual property interests of sponsors, and ultimately, improving patient care. In addition, it is

important to establish an independent review panel to review all requests for access to the shared data and to make access to trial data transparent. Even though not perfect, these steps can substantially reduce the risk of privacy breaching, and can set professional expectations and standards for responsible behavior.

For example, Project Data Sphere [5], an independent not-for-profit initiative of the CEO Roundtable on Cancer's Life Sciences Consortium, operates a platform with a free digital library-laboratory that provides the research community with the capability to broadly share, integrate, and analyze historical, patient-level data from academic and industry phase III cancer clinical trials. The consortium's goal is to help the cancer community unlock the potential of valuable data by generating new insights and opening up a new world of research possibilities.

Request from the Regulatory Agencies and Journal Editors

In the past few years, the European Medicines Agency (EMA) had published a series of guidance to instruct pharmaceutical companies to prepare de-identified clinical reports and clinical data for studies submitted to EMA for marketing approval. The agency had informed sponsors that these data will eventually be published openly on their websites.

In their guidance [6], the EMA had provided instructions for the anonymization of clinical reports for the purpose of publication in accordance with EMA Policy-0070, and how best to anonymize data in accordance with the legal framework and available standards. Similar guidance was also provided in Appendix B of the IOM document for sharing clinical trial data.

Following the IOM's initiative, many Independent Review Boards of study protocols have started to ask study sponsors about their data sharing plan and their approaches for the protection of participants' privacy as one of the considerations for them to approve a proposed study. One can reasonably expect this kind of practice to become a norm for future clinical trial protocol reviews.

In January 2016, the International Committee of Medical Journal Editors (ICMJE) published an article [7] to advocate the necessity of clinical trial data sharing for manuscripts to be published in their journals. The ICMJE believes that there is an ethical obligation to responsibly share data generated by interventional clinical trials, as clinical trial participants have put themselves at risk, and in a growing consensus, many funders around the world, including foundations, government agencies, and industry, now mandate data sharing. As a consideration for publication of a clinical trial manuscript in the member journals, the ICMJE proposes to require authors to share with others the de-identified individual-patient data underlying the results presented in the article soon after publication. The ICMJE also proposes to require that authors include a plan for data sharing as a component of clinical trial registration.

Challenges and Caveat

Even though clinical data sharing is gaining momentum with great benefits in enhancing medical research for the general public, one should not be overly optimistic in setting high expectations. The willingness to share data from for-profit organizations such as pharmaceutical companies can be a concern due to the protection of commercially sensitive information. Any loss of protection on their confidential commercial information may financially penalize them substantially. The legal issues surrounding the invasion of privacy of trial participants is another reason for the reluctance to share data, since no method is foolproof even after careful de-identification of personal information.

Statistical data analysis can post another challenge. Since releasing original data to the public is highly unlikely, data will most likely need to be modified through some form of anonymization procedure. Depending on the extent of anonymization performed on the data, information in the original database will inevitably be compromised, making replication of the original results difficult, if not impossible. The significance of covariates on the treatment effect in the original model may be altered due to the loss of precision because of modifications to the original data. Such alterations could subsequently generate misleading conclusions to the research findings. In addition, if multiple databases for the same disease class were contributed by several organizations, the data may not be readily pooled for meta-analysis due to the potentially different anonymization procedures performed on the original databases. Therefore, the significance of the effect of covariates on the treatment effect will truly be data dependent. IOM and EMA both advocated a good balance between data de-identification to protect patient privacy and high level data utilization for secondary research; however, it is not easy to determine how this balance is to be achieved.

To make data easily accessible, a user-friendly platform to host the collection of data and clear instructions for accessing data are critical. Since the raw data is most likely not downloadable by external researchers, the host of the platform needs to either provide the varieties of software for data analysis or restrict users to a certain number of software applications available, and provide a secure channel with a high-level of security control for researchers to login to fulfill analysis needs.

Another concern is the qualifications of researchers who want to utilize databases to conduct data analyses. Even professional statisticians may not always produce good quality analytical results, let alone individuals without proper training and experience. Incomplete analytical results with potentially biased findings can produce more harm than good. Therefore, some well-defined reviewing processes with sufficient quality control from neutral organizations are necessary since the research findings, originally intended for advancing medical research, could possibly be published, used for financial gains, or other unintended purposes.

References

1. Institute of Medicine (2015) Sharing clinical trial data: maximizing benefits, minimizing risk. Washington, DC: National Academies Press.
2. Health Human Services (2012) Guidance regarding methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. Washington, DC: HHS.
3. Emam KE (2013) Guide to the De-Identification of Personal Health Information. CRC Press.
4. Damian Garde (2014) TransCelerate BioPharma Inc. Releases Recommended Approach for Protecting Personal Data in Clinical Study Reports.
5. Project Data Sphere
6. European Medicines Agency (2016) External guidance on the implementation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human use.

7. Backus J, Baethge C, Bauchner H, de Leeuw PW, Drazen JM, et al. (2016) Sharing Clinical Trial Data - A Proposal from the International Committee of Medical Journal Editors. Ann Intern Med 164: 505-506.