

DATA SHARING

Power to the People: Participant Ownership of Clinical Trial Data

Sharon F. Terry^{1*} and Patrick F. Terry²

Participation in clinical trials is dismally low. In this age of electronic sharing of information of all sorts, trial participants can easily share clinical trial data. The benefits of participant ownership and sharing of trial data appear to outweigh the risks. Thus, the time has come to crowd-source data for diagnostic and therapy development.

COMMONS CONCEPT IN THE CLINIC

“I think that many ... academic researchers still are working in a hunter-gatherer society—not just hunter, but hunter-gatherer—and actually feel as if you pay them to generate data, that they have the right to mine it, they have the right to keep it. ... You should think about why—and this is what patients come up to us and say—why can't you guys share more?” (1)

This question is one that should be reverberating throughout every community of patients, scientists, and policy-makers: Why can't those who contribute data to clinical studies—the participants—decide where the data will reside and who will be able to use it? What if the commons concept—that resources of various kinds are owned collectively—was applied to clinical trial data, with participants leading the way in creating dynamic interfaces for data sharing (Fig. 1)? Would we wake up to a brave new world or to devastation wreaked by havoc?

At the present time, clinical trial data reside with the sponsor of the trial, which is usually a company or an academic institution. A great deal has been written about the failure of the clinical trial system as it currently operates (2–4). However, conversations about changes in data ownership policies are occurring, and some experiments are actually under way in which data control has shifted to the person contributing the data. The ability to move data—the liquidity—is increasing overall, in particular through the potential availability of electronic medical records (EMRs) and personal health records (PHRs). It stands to reason that clinical trial data will become liquid as well.

Data liquidity, coupled with standards and interoperability, is usually an uncontroversial goal. Hospitals, clinics, physicians, health insurers, and pharmaceutical and biotech companies are all interested in simpler data flow so that they can deliver more efficient services. However, an interesting set of challenges and opportunities emerges when clinical trial data are in the hands of the individuals who contribute the data. Participant ownership and sharing of clinical trial data raises conundrums throughout the entire system, including the fundamentals of trial design, intellectual property, protocol management, and regulatory oversight.

SHARING: MORE THAN GOOD MANNERS

Why would clinical trial participants share data? A trivial reason might be because the means exist to do so. Participants can control data with easy-to-access and easy-to-use tools. In the age of Facebook and Twitter, individuals share a great deal of information

much more easily and willingly than in prior times; these same people might be inclined to share clinical trial data as well. However, this reason dismisses the impending possibility too trivially.

Many individuals who are engaged in clinical trials place a desperate hope in a very constrained and limited attempt to mitigate manifestations of a life-threatening, or at least poorly treated or untreatable, illness. They place themselves in a clinical trial with tremendous hope. If they come to the end of a trial without the result they sought (mitigation of disease, better quality of life, or the ever elusive and improbable “cure”), they are easily motivated to share the data they contributed to the trial. With one hope dashed, they seek another. This is particularly true for conditions for which there is no efficacious treatment, such as Parkinson's disease. One has to simply examine the phenomenon taking place in the various PatientsLikeMe (5) Web-based communities to gain a glimpse of what a world of shared patient data looks like. Daily entries by tens of thousands of individuals indicate the drive some people possess for sharing data with others.

LIMITATIONS INCENTIVIZE SHARING

For any diagnostic test or therapy to be approved, human trials must be conducted. However, too few individuals engage in trials, and 50% of clinical trial sites enroll one or no patients in their trial (6). A major blockade in the development of new tests and treatments is a lack of clinical trial participants. The current system for recruiting individuals relies on antiquated means

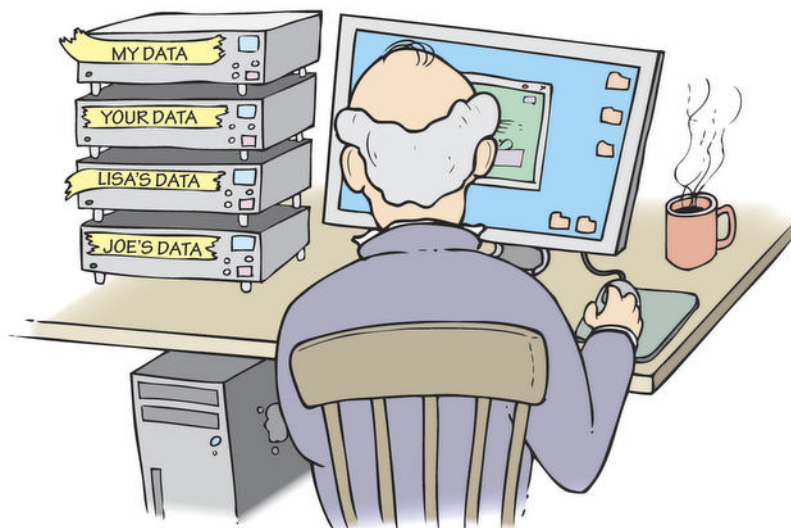


Fig. 1. “What’s my data is mine and what’s your data is also mine.”—Sydney Brenner, on data-mining (18).

¹President and CEO, Genetic Alliance, Washington, DC 20008, USA. ²Principal, Scientia Advisors, Cambridge, MA 02142, USA.

*Corresponding author. E-mail: sterry@geneticalliance.org

of contacting and engaging individuals. It may be that these antiquated practices result from the lack of resources remaining after the many burdensome and costly aspects of designing and obtaining approval for the trial protocol are finished.

Clinical trial protocols suffer because an enormous part of the work in establishing the trial is whittling down the ideal protocol to a manageable number of elements that will satisfy regulatory authorities. Therefore, the trial sponsors must discard a great deal of potential data fields. If participants own their data, then they might be very interested in collecting a greater number of elements. Real-world conditions are not “allowed” in clinical trial design, and it is possible that the constrained data sets can constitute the proverbial “garbage in”: nearly perfect data that are so narrow in scope that they are uninformative. Trial conditions necessarily create an artificial scenario within which data are managed in order to facilitate the observation of a statistically meaningful effect if one indeed exists. The constrained scenario and statistical narrowing are necessary for regulatory approval, but these conditions sacrifice the heterogeneity necessary for personalized medicine.

The advent of personalized medicine, effectively precision-stratified medicine, represents continuums that need biomarkers along the way to enable cluster analysis of patient populations. These requirements bend the usual clinical trial rules to the point of breaking. For example, if a clinical study doesn't demonstrate the desired therapeutic response to a drug, then observational studies might be beneficial to determine whether a selected subgroup of patients within the trial population exists for which the drug is effective. However, protocols usually are not flexible enough to allow for the adaptive clinical trials necessary to obtain a range of answers about the effectiveness or appropriate use of a drug that is undergoing testing. Studies of agents that prevent disease suffer even greater constraints, in part because commercial markets will not develop most preventative treatments. For example, companies can't benefit as handsomely from commercializing a functional food or lifestyle change.

Furthermore, as clinical trial participants report their experiences in virtual groups such as listservs and Facebook, these patients influence one another and the trial in real time. These modern trends and technology are making randomized trials more

and more difficult to carry out. Thus, these real-life limitations in the current clinical trial system are creating social and scientific pressures to aggregate data in novel ways.

POTENTIAL UNLEASHED

If many individuals could share their clinical trial data themselves, either through biometric tracking mechanisms, EMRs and PHRs, or through questionnaires, the research world would be a drastically different place. These “citizen scientists” could radically overhaul an ineffective system.

As stated above, the intensely controlled environment of clinical trials creates an artificial universe. It is clear that introducing the dynamics of the real world creates data integrity problems and that sharing these data and comparing them in numerous ways would do the same. However, because informatics is able to address the issues inherent in collecting and analyzing seemingly limitless data points, it appears likely that great strides in clinical research will be made through enhanced data sharing. It is quite clear that environment, compliance with therapies, and even less-tangible parameters such as quality of life play a large role in health and the effectiveness of treatments (7). Patient control of clinical data will introduce both chaos and richness into the system.

Models of participant control of clinical trial data that are currently being shaped and tested give us some sense of what the world of clinical data sharing can and will entail. There are many models that differ from one another in how data are managed and shared, at which points they are shared, and what is embargoed.

In simple examples of data sharing by trial participants, the Internet is rife with commentary from individuals who report the various measures they can remember from their trial data. Although this is not clinical trial data ownership in the strictest sense, it formed an early entrée into the world of sharing parameters and experiences while involved in a clinical trial. In some instances, it is possible for participants to observe effects of drugs and placebos before the trialists do, through these online comparisons.

The small, privately held company Private Access gives individuals the potential to enter data that could be useful for clinical trial recruitment purposes. Privacy controls on the data are set by individuals according to their preferences, allowing them to determine whether all scientists can see all of

their data, select scientists can see their data, or no scientists can see their data. In the future, these controls would open access to an individual's data held by medical professionals, labs, and other places in which these data exist. Several pharmaceutical firms have taken heed of this foray into patient access and control, and Pfizer has entered into a collaboration to advance this functionality with Private Access (8).

Another initiative is Genomera. This Silicon Valley-based company not only gives individuals a platform through which to share genomic and phenotypic information but also provides them with the tools to create and participate in their own clinical trials. In these novel Genomera-hosted studies, participants decide to subject themselves to an observation or intervention protocol and report the results back to the group for joint analysis. Distributing the locus of control in this manner allows participants to opt in to studies as they desire and frees study organizers from the legal and regulatory burden of experimenting on others.

Providing information and technology to create trials will allow all sorts of questions to be asked, some to be answered, and even some to move on to application. The last step is a tricky one, and Greg Biggers, Genomera's founder, thinks that U.S. Food and Drug Administration oversight, if informed, could eventually provide helpful guidance to trial participants and beneficial protection to prospective users of the conclusions (9). In the absence of new regulatory models, Genomera captures expertise among scientists and participants, providing transparent access to opinions about risks, ethics, and efficacies of the hosted studies.

Software company 5AM Solutions released a third-party add-on for the Firefox browser called SNPTips, which, after being associated with your 23andMe genome sequence data, trolls the Web pages you are viewing for SNPs that are relevant to your genome. For example, when reading a news story or journal article online that discusses the impact of various SNPs, those SNPs are highlighted. If you click on them, your genotype pops up attached to the SNP. This saves the hardcore citizen scientist from manually searching for the SNP of interest in raw sequence data.

Sage Bionetworks and Genetic Alliance have worked together to create the Clinical Trial Comparator Arm Partnership (CTCAP). Still in its first year, the CTCAP includes seven large companies, a clinical

trial research organization, and a pharmacy benefits management company, which have all committed one or more data sets to the commons, including clinical and genomic data. In this case, the participants are releasing their data through the agreements that they have made with the companies during the consent process. All in all, clinical and genomic data from more than 20,000 patients will be made available for building better maps of disease in an open-access system. Prospectively, if companies such as Pfizer use the Private Access system for enrolling participants, then those participants can make specific decisions, such as the one to contribute to CTCAP.

Participants also contribute data to the nonresponders project at Sage Bionetworks led by Stephen Friend and Rich Schilsky. This is a completely open project through which a cohort study, set to begin in 2011, will analyze samples from responders and nonresponders to approved cancer drugs. Its goal is to identify those patients who are so unlikely to respond to the approved first-line drugs for specific tumor types that they could bypass those treatments and instead be given investigational drugs up front before prior therapy. All of the clinical and genomic data will be made available for building new disease maps (10).

Although these are all meaningful and exciting forays into the world of participant ownership of data, they entail fairly small numbers of individuals relative to the power that is needed for a revolution in clinical trial design. These efforts are just the tip of the iceberg, and, much as we have seen in informatics in other industries, there will be a surge in the availability of tools to participate in trial data sharing. Individuals will enter data into third-party software solutions for comparing, contrasting, correlating, and analyzing their data. With a critical mass, the tools themselves will improve, allowing a crowdsourcing effect, similar to the improvement in quality of Wikipedia as more individuals participate. Further, if success can be demonstrated by the model initiatives described above, some of the criticism of lay participation in clinical research would abate. Most traditional scientists disparage this movement, declaring that no useful findings will emerge from the efforts of lay people who do not know how to analyze and interpret information from clinical trials.

Individuals already participate in rich data collection efforts. Some examples include global positioning device applications

on smartphones such as MapMyRide, or Nike sensors in running shoes that measure activity, or a barcode reader on the iPhone that scans food items and logs in nutrients. One such scanner is coupled with the DailyBurn Web site to log an individual's food intake, activity, and sleep. One could imagine that some of the emerging models might eventually tie into these systems and, with the participants' permission, acquire environmental and life-style data (after all, our smartphones can track where we are at all times). This might give individuals an experience of active engagement in clinical trials. The results coming from the analysis of these data, particularly if it is done in real time, might make individuals feel more empowered when it comes to participating in clinical trials.

One could even imagine that compliance studies that use data from scanning a prescription bottle—or even better, an electronic pill container—and then provide simple mechanisms for post-approval studies and adverse-event reporting could be far more effective than today's typical studies. Young people might be enticed to participate in data collection through games—more than 80 million people play Farmville on Facebook, which, at 600 million participants, is the third-largest “country” on the planet, after China and India. There are some efforts under way to make type 1 diabetes management fun and relevant to young people through gaming. These interactive engagements are here to stay, and if clinical trials were interactive, there would undoubtedly be a much higher level of engagement than the currently estimated less than 3 to 5% (2, 6).

WHAT TO DO WITH ALL THE DATA?

Hypothesis-generating studies. If participants controlled clinical trial data, then the strict definition of data as linked observations that are gathered to test a hypothesis (11) would be difficult to control. Hypothesis-generating trials would arise in abundance. The usual criticism of such data sets is that they are not useful for generating hypotheses because the data compared were collected for different purposes and thus are not informative. However, Web sites such as PatientsLikeMe and 23andMe have shown through research performed in their communities that there is power enough to replicate much more expensive studies (12, 13).

As examples of translational results obtained from mining multiple clinical data

sets, three recent papers by members of Sage Bionetworks report the identification of key network modules that drive disease (14–16).

Another argument against citizen scientist data assembly declares that the costs of collecting such data will be prohibitively high (17). This argument has become less and less plausible with the phenomenal decrease in the costs of data collection and even more automated data cleaning and winnowing software.

Although the data that an individual contributes to clinical trials can be considered theirs without much of a stretch, there is concern that data that are generated are proprietary to the sponsor given the effort and expense they invested to collect it. This is similar to the argument of scientific journals at the dawn of the public access movement, when they determined that it was one thing to consider the article the property of the author and quite another to determine the same about the article once it was edited. There are varying opinions about this matter; however, at this time all NIH-funded peer-reviewed articles must be deposited in PubMed Central within 12 months of the official date of publication. It follows then that clinical data that are generated with public funding also should be shared for the good of the public health.

Concerns about privacy are always a consideration when sharing clinical data. If participants control the distribution of their data, then there are fewer concerns because the major privacy law in the United States, the Health Insurance Portability and Accountability Act (HIPAA), does not apply; individuals are free to share their own clinical information under HIPAA. However, privacy systems with granular controls are desirable because individuals will want to share information selectively, as is their right. Further, the various options will undoubtedly be complex, and guides such as those used by Private Access will be essential.

SOMEDAY HAS ARRIVED

Fantastic tools exist for adaptive clinical trial designs, crossover studies (ones in which patients receive a series of sequential treatments), behavior monitoring, patient-centric risk-benefit threshold measurements, and post-approval data collection. Like sausage-making and legislation-passing, it would be disturbing for healthy controls to see the mechanisms and bureaucracies associated with the internal workings of the clinical trial processes. The time is ripe for

individuals to become active participants in clinical trials, to create a movement to build the commons with such data, and to actively share them.

REFERENCES AND NOTES

1. S. Friend, Canadian scientists must promote a culture of data sharing to promote innovation. Paper presented at Capturing Innovation for the Public Good, 15 September 2010, Toronto, Canada. www.gairdner.org/2009events/capturinginnovationforthepublicgoodreport.
2. National Cancer Institute, Doctors, patients face different barriers to clinical trials (2010). www.cancer.gov/clinicaltrials/developments/doctors-barriers0401.
3. M. A. Hamburg, F. S. Collins, The path to personalized medicine. *N. Engl. J. Med.* **363**, 301–304 (2010).
4. R. English, Y. Lebovitz, R. Griffin, *Transforming Clinical Research in the United States* (Institute of Medicine of the National Academies, Washington, DC, 2010). www.iom.edu/Reports/2010/Transforming-Clinical-Research-in-the-United-States.aspx.
5. PatientsLikeMe, (2010). www.patientslikeme.com.
6. CISC RP, Professionals: Clinical Trial Facts and Figures (2010). http://www.ciscrp.org/professional/facts_pat.html#5.
7. L. A. Moyé, Disciplined analyses in clinical trials: The dark heart of the matter. *Stat. Methods Med. Res.* **17**, 253–264 (2008).
8. Ten questions for Private Access. *Forbes* (2009), www.forbes.com/2009/09/08/private-access-ten-questions-entrepreneurs-promising.html.
9. E. Dolgin, Personalized investigation. *Nat. Med.* **16**, 953–955 (2010).
10. S. Friend, Cohort study for nonresponders: Patients donate their samples. Personal communication to S. F. Terry, Washington, DC, USA (2010). Used with permission.
11. J. M. Drazen, Who owns the data in a clinical trial? *Sci. Eng. Ethics* **8**, 407–411 (2002).
12. N. Eriksson, J. M. Macpherson, J. Y. Tung, L. S. Hon, B. Naughton, S. Saxonov, L. Avey, A. Wojcicki, I. Pe'er, J. Mountain, Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genet.* **6**, e1000993 (2010).
13. P. Wicks, M. Massagli, J. Frost, C. Brownstein, S. Okun, T. Vaughan, R. Bradley, J. Heywood, Sharing health data for better outcomes on PatientsLikeMe. *J. Med. Internet Res.* **12**, e19 (2010).
14. V. Emilsson, G. Thorleifsson, B. Zhang, A. S. Leonardson, F. Zink, J. Zhu, S. Carlson, A. Helgason, G. B. Walters, S. Gunnarsdottir, M. Mouy, V. Steinthorsdottir, G. H. Eiriksdottir, G. Bjornsdottir, I. Reynisdottir, D. Gudbjartsson, A. Helgadóttir, A. Jonasdottir, A. Jonasdottir, U. Styrkarsdottir, S. Gretarsdottir, K. P. Magnusson, H. Stefansson, R. Fossdal, K. Kristjansson, H. G. Gislason, T. Stefansson, B. G. Leifsson, U. Thorsteinsdottir, J. R. Lamb, J. R. Gulcher, M. L. Reitman, A. Kong, E. E. Schadt, K. Stefansson, Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428 (2008).
15. Y. Chen, J. Zhu, P. Y. Lum, X. Yang, S. Pinto, D. J. MacNeil, C. Zhang, J. Lamb, S. Edwards, S. K. Sieberts, A. Leonardson, L. W. Castellani, S. Wang, M.-F. Champy, B. Zhang, V. Emilsson, S. Doss, A. Ghazalpour, S. Horvath, T. A. Drake, A. J. Lusis, E. E. Schadt, Variations in DNA elucidate molecular networks that cause disease. *Nature* **452**, 429–435 (2008).
16. X. Yang, J. L. Deignan, H. Qi, J. Zhu, S. Qian, J. Zhong, G. Torosyan, S. Majid, B. Falkard, R. R. Kleinhanz, J. Karlsson, L. W. Castellani, S. Mumick, K. Wang, T. Xie, M. Coon, C. Zhang, D. Estrada-Smith, C. R. Farber, S. S. Wang, A. van Nas, A. Ghazalpour, B. Zhang, D. J. Macneil, J. R. Lamb, K. M. Dipple, M. L. Reitman, M. Mehrabian, P. Y. Lum, E. E. Schadt, A. J. Lusis, T. A. Drake, Validation of candidate causal genes for obesity that affect shared metabolic pathways and networks. *Nat. Genet.* **41**, 415–423 (2009).
17. M. A. Rodwin, The case for public ownership of patient data. *JAMA* **302**, 86–88 (2009).
18. S. Brenner, The worm's turn. *Curr. Biol.* **12**, R713 (2002).
19. S.F.T. is an informal advisor to Private Access and is partnering with S. Friend on the CTCAP. We are grateful to all who participate in clinical trials, recognizing that many times the trail they blaze does not benefit them but benefits those who come after them, and to S. Friend of Sage Bionetworks and J. Wilbanks of Creative Commons for their leadership and companionship on this journey toward a greater commons. The openness experiment in which the council and staff of Genetic Alliance have engaged has greatly influenced our thoughts and opinions and is a wonderful gift.

10.1126/scitranslmed.3001857

Citation: S. F. Terry, P. F. Terry, Power to the people: Participant ownership of clinical trial data. *Sci. Transl. Med.* **3**, 69cm3 (2011).