Toward Clarity on Transparency

An evolution in thinking - and action - about sharing patient-level clinical trial data

Insights From the Fourth Installment of the Clinical Trial Data Transparency Forum
Contents

Why, and Why Now? ......................................................... 1
Challenges and Risks ..................................................... 2
Sharing What Works ...................................................... 3
  Implement a structured review process. ......................... 3
  Clearly define what’s in and out of scope. .................... 4
  Require a statistical analysis plan. ............................... 4
  Encourage new science. .............................................. 4
  Don’t be paralyzed by potential risks. ......................... 5
  Don’t try to be everyone’s center of excellence. ............ 5
  Collaborate in broad context. ...................................... 5
  Dedicate the people. .................................................. 5
  Share best practices. .................................................. 5
  Do something. Do it now. .......................................... 5
A Timely Announcement .................................................. 6
Closing Thoughts .......................................................... 6

View on-demand recordings of all the presentations from the fourth Clinical Trial Data Transparency Forum at sas.com/ctdt4.

Forum attendees represented AstraZeneca, Bristol-Myers Squibb, Celgene Corp., Duke Clinical Research Institute, Duke University Health System Inc., Eisai Inc., Eli Lilly, FasterCures, Gilead Sciences, GlaxoSmithKline, Janssen R&D, Johnson & Johnson, MedData, Medidata Solutions, MedImmune LLC, Merck, NCHICA, Paarlberg & Associates LLC, Pfizer, Project Data Sphere LLC, Quintiles, Sanofi, Takeda, UCB Biosciences GmbH, University of Pittsburgh. Comments in this article represent a compendium of general discussion at the forum and not the opinion of any particular organization.
“Access to the underlying (patient level) data that are collected in clinical trials provides opportunities to conduct further research that can help advance medical science or improve patient care. This helps ensure the data provided by research participants are used to maximum effect in the creation of knowledge and understanding.”

That’s the word from ClinicalStudyDataRequest.com, the data-sharing consortium that represents the founding sponsor, GlaxoSmithKline, and now nine other sponsors.

Sharing patient-level data for altruism and public good? The ideal was not initially embraced.

“When I think back two years to when GSK first started talking about this, there was certainly some concern among staff,” said Paul McSorley of GlaxoSmithKline, a pioneer in the data-sharing movement. “How much work is this going to be? What happens if or when researchers reach different conclusions from our own? We’re way past that now. GSK scientists recognize that these concerns, while real, can be mitigated – and because there is so much support for the potential value data sharing can bring to the medical community, we are very proud of what GSK is doing here.”

In the year since SAS hosted the first Clinical Trial Data Transparency Forum, we have seen a notable shift in organizational cultures and the tenor of the discussions:

• Stage 1. “We see merit in the idea, but we also see many ways it could go wrong.”
• Stage 2. “We need to do something before external entities impose a data-sharing framework on us.”
• Stage 3. “We’re excited to be at the forefront of creating policies and processes to make this work.”
• Stage 4. “This may not be the final state of things, but here’s what has been working for us.”

The fourth forum, held at SAS in Cary, NC, on Oct. 2, 2014, exemplified stages three and four: endorsement for data sharing and more tangible progress to show for it. These forum events are designed to facilitate and formalize the conversations that have been taking place in various corners of the industry and academia.

Why, and Why Now?

The life sciences industry seems to have come to agreement that greater access to patient-level data is a good thing – good for science, good for business and good for humanity. Data sharing can lead to discovery of new trends and associations that generate new insights or hypotheses for further research. Data sharing enables objective, third-party review and validation of study results, thereby building public trust. Data sharing honors the valuable information provided by patients and researchers in previous clinical trials and extends the future value of their efforts.

“It’s incredible data,” said Dr. Stephen Freedland, Associate Professor of Surgery and Associate Professor in Pathology, Duke Urology, at the Duke University School of Medicine. “You guys spend millions of dollars gathering the data; the data is clean, there are no holes, and it’s prospective versus retrospective. So the richness of the data is incredible. From an academic perspective, it allows PhD access to patient data to test new ideas, get quick answers and validate prior studies. For example, we had seen that obesity is a risk factor for prostate cancer; when we saw that in other data sets, this adds another point of validation.”

Data sharing also enables broader research on a disease or condition rather than on the efficacy of a product. Take depression, for example. “If it were possible, it would be really great to take all of the depression studies that have ever been done, and find things that were not captured in routine measures,” said Dr. Ronald Krall, Adjunct Associate Professor in the Department of Neurology and Center for Bioethics and Health Law at the
University of Pittsburgh (and former Chief Medical Officer at GlaxoSmithKline). For example, is there a more sensitive indicator of efficacy of an antidepressant? How do the various statistical methods of dealing with dropouts perform? Which methodological approaches work best? “These insights might have been captured, but they’re not made evident in the way we analyze the data,” said Krall.

Bottom line, the public expects it of us, said Stephen Weitzman of MedData. “Because of the technology advances of the last 10 years, breakthroughs in genomics, supercomputing and big data analytics for health care, there’s a public expectation that we can handle big data. That’s a very important factor in the public being willing to share their data and help resolve some of the problems.”

Challenges and Risks

For all the benefits, there are caveats as well. Unless appropriate safeguards are in place, open access to clinical trial data could compromise patient privacy, enable faulty science, and be a resource-intensive burden for trial sponsors and data stewards. These are all outcomes the industry wants to avoid before going too far down this path. These common objections are understandable.

Why should we take on this extra work?

Under the publish-or-perish dictum of the academic research environment, people often prefer to hoard their data. They fear if they share their data with a big consortium, 100 other named authors will get published on it. In the business environment, the fear is economic loss. Will we be scooped or discredited?

What if we don’t have the requested data in electronic form?

Want data from a study performed in the 1990s? It may not be available in electronic form, said Ben Rotz, Medical Transparency Adviser at Eli Lilly. “Data in paper format would be tough to access and cost a significant amount of money to electronically code.” Where data from multiple sponsors is available electronically, it can still be difficult to get it into one analytical environment due to variances in coding, standards and language. “If you want to bring trials together, you’re going to have real problems making sure that what was recorded 15 years ago is the same format as what’s being recorded today.”

What if secondary researchers get it wrong?

Two people can analyze the same data and come up with very different results. A study of medical literature from 1966 to present found that of 37 published studies that reanalyzed randomized clinical trials, fully one-third of them came to different conclusions about whether the intervention worked. 1 Who’s right?

“The fear is that people are going to take the data out of context,” said Marla Jo Brickman, PhD, Senior Director/Team Leader of the Clinical Trial Disclosure Group at Pfizer. However, this concern does not seem to be as big as previously thought, Brickman noted. “One failed study does not a failed product make. And most of the data requests we have received were not to validate our findings but actually to expand on the original learnings.”

What if the data is intentionally misused?

Data sharing is a double-edged sword, said Dr. Eric Peterson, Executive Director of the Duke Clinical Research Institute. “On the one hand, there is the ability to open up the world, to find new things that you could have never afforded to do yourself or that someone else’s creative thoughts will help open up – answers that were not heretofore available. On the other hand, there are competitors that might want to look at your data to find some competitive advantage for themselves, or worse, change the marketplace impression of your product because they got their group of investigators to reanalyze your data and put it in a different light.

“You can ask a question a zillion times [until you get the answer you want]. I can change the primary endpoints slightly, change the population, add a few more variables, run it again and again and again – and sooner or later I will either get the answer I want or something that seems somewhat interesting, and publish on it. That is not science.” Unfortunately, the contrarian perspective – which may be skewed from the truth – is more likely to make it to the media or a journal. Consider the alternative headlines: “Popular Drug Works as Originally Thought.” That’s not a story. “Drug Doesn’t Work or May Harm You.” Now that’s a story.

If notoriety or publicity are the researcher’s aims, data dredging can get to that end. But is that a prevalent risk? “Someone could spin the data to make the product/company look bad,” said Freedland. “But in general, most people want to move science forward and will use the data for good.”

1 Shanil Ebrahim, PhD; Zahra N. Sohani, MSc; Luis Montoya, DDS; Arnav Agarwal, BSc; Kristian Thorlund, PhD; Edward J. Mills, PhD; John P. A. Ioannidis, MD, DSc, “Reanalyses of Randomized Clinical Trial Data,” Journal of the American Medical Association (JAMA), Sept. 10, 2014
Where will we find the resources to respond to data requests?

It’s not a trivial effort. After the data set and documents have been located, it can take an average of five to seven days of effort depending on the scope of the proposal, said McSorley.

That includes one to three days for document redaction to remove personal information (less if there’s no clinical study report); three days to anonymize the data to protect patient privacy; and a day to check the proposal and load the data into the access system, applying quality control processes along the way. The process could be twice as long for a large or historical study with complex requirements and multiple editions of the data, or if the researcher has lots of questions about the data.

“We hope to bring those numbers down, especially the anonymization piece, through automation,” said McSorley. “But we want to be sure we’re doing it right.”

Data sharing is not zero risk, but nothing in life is zero risk. The risk of not sharing the data – all the great science lost – all the opportunity costs, all the dollars required to create other cohorts, dollars that are not being spent examining new questions – all of these are greater than the risk of sharing.”

Dr. Eric Peterson, Executive Director, Duke Clinical Research Institute

Another option is the Yale University Open Data Access (YODA) model, where two independent research groups review and analyze the data request, under guidance from an independent steering committee of clinical research and biomedical ethics experts and an advisory committee of clinical practice leaders.

In Pfizer’s model, data requests get a preliminary triage to determine whether they are feasible – and in or out of scope – and then the request is sent to an internal Pfizer review panel. This panel includes people with the requisite regulatory, medical, statistical and legal understanding of the data. If the request is declined outright or in part, then it goes to Pfizer’s independent external panel, whose decision is binding.

“To date we have not had to send anything to the external panel, because everything that has been in scope is either pending or has been approved internally,” said Brickman. “There may have been some perceptions that a company can’t review its own data requests because they’re not going to give folks access to it. That has not been the case for us. We have been making the data available, and we have not declined any [in-scope] requests.”

Brickman noted cases where having an internal review panel was actually beneficial because these people knew the requested data had already been published somewhere and were able to provide something the researcher did not know about and therefore did not have to replicate.

How do you ensure data access by responsible researchers who have knowledge of good analytical and statistical principles and the science behind the trial? To meet this need, there is wholesale endorsement for using independent review panels to review research proposals and either approve or deny the associated data access requests.

Sharing What Works

Participants in our October 2014 forum shared their experiences on what has worked for their organizations.

Implement a structured review process.

Many of the concerns expressed earlier – validity of the research, protection of patient privacy, possible misuse of the data for competitor gain or headline grabbing, and protection of data exclusivity for the primary author/investigator – can be addressed by having a solid process for reviewing and approving data requests.

The review process could follow a hybrid model, where requests go first to a review panel of the original study sponsor, with only denied requests escalated to an external panel.
Clearly define what’s in and out of scope.

“We’re trying to be as visible and transparent as possible as to what we can and cannot provide, and what is required to review a request,” said Judy Bryson, PharmD, Senior Clinical Program Director in the Immunology Practice at UCB Biosciences Inc. “No organization can promise an open book. It’s just not feasible.” You have to state limits, protect the privacy of patients with rare conditions (who could be easily identified even from anonymized data), and protect commercially confidential information.

“We are very clear in our policy about what’s in scope and what’s out of scope,” said Brickman of Pfizer. “When you have a policy in place to vet all requests, you become above scrutiny, because it’s clear we’re not favoring an individual or an institution.” If a request is denied, there’s no perception of unwillingness to share, just adherence to publicly stated policy.

How far back should study data be made available, and when? Even members of our panel discussion, whose companies participate in one data-sharing program, differ on this count. Pfizer’s policy states that studies concluded in 2007 or later are fair game. There’s a waiting period of two years from the time a study is completed to the time the data will be made available, so the internal team and investigators have their opportunity to do secondary analysis and publish. Other companies set different threshold dates, generally based on the availability of electronic data.

Require a statistical analysis plan.

“We don’t want to give data to people who just want to fish,” said Brickman. “We are asking for bona fide requests for scientific research, so we do require a statistical analysis plan, and we do require a statistician to be a part of the team. We’ve had a couple of requests withdrawn because they’ve refused to give us the statistical analysis plan until we told them they could have the data.”

“I’m a huge believer in a statistical analysis plan as the way to improve primary and secondary research,” said Peterson. “A prospective plan can avoid spurious findings.”

Suppose you specify in your statistical analysis plan what the question is and how you mean to ask it. As the research unfolds, can you now deviate from that? Absolutely, says Peterson. “There are reasons you might change the questions you’re asking based on the answers you are getting along the way. But if the plan and its evolution are transparent, readers know the degree to which this research process is hypothesis-generating versus hypothesis-concluding.”

Encourage new science.

Few research proposals seek to replicate previous research, Rotz reported. The overwhelming number are quests to create new science. McSorley concurred, noting that only one in 23 data requests was a proposal to confirm results of a previous GSK study.

Freedland provided a compelling example of the value for creating new science. Starting with “academic” data, he developed retrospective cohorts within the Veterans Administration system and various universities. This research generated some interesting ideas and resulted in published, hypothesis-generating papers about prostate cancer. When multiple studies started saying the same thing, Freedland wanted to look closer and test these discoveries in a clinical trial.

Looking at it as an epidemiological cohort, the study would follow patients for multiple years, looking at baseline characteristics to see who developed prostate cancer of what degree. “I sought access to pharma data to validate my ideas,” said Freedland. “I didn’t really care about the original trial per se – did drug A do better than drug B – but rather viewed the data as a prospective cohort study for secondary analysis.”

Freedland met an MD at GlaxoSmithKline who was very open about sharing data. After hearing Freedland’s ideas, he offered full access to the data — and even a small grant to analyze it. That collaboration has produced 12 research papers, none of them related to reanalyzing the efficacy of the drug, but rather looking at risk factors. The result was clear evidence that smoking is correlated with more aggressive prostate cancer, and obesity is also a significant risk factor. Those insights were revealed by making new use of pre-existing data, just one example of the potential of well-managed data sharing.

Our panelists agreed that shared data shouldn’t be used only to question the validity of previous primary research — and few requests are of this type. The overwhelming majority of requests are intended to create new science.
Don’t be paralyzed by potential risks.

Might secondary research turn up a skeleton you didn’t know was there? As Krall noted, you might be afraid of learning something you didn’t want to know, or something that could harm your product or competitive position, but if you’re committed to knowing everything you can possibly know about your products – and you’d rather know it sooner than later – transparency is the ticket.

“Nothing in life has zero risk,” said Freedland. “Everything in life has risks; the goal is not to eliminate risk. The goal is to make smart choices when we’re willing to accept that risk because the benefits are high enough.”

Don’t try to be everyone’s center of excellence.

The original study sponsors will probably answer questions about the data to help researchers in their work, but they shouldn’t be expected to give advice about scientific methods or how to build statistical models. Data requesters should bring those skills with them.

Collaborate in broad context.

Bring together diverse disciplines. Project Data Sphere is a great example. Dr. Kald Abdallah, Chief Project Data Sphere Officer, said the organization has a threefold mission: Grow the data, expand and stimulate use of the data, and nurture the network of collaboration to support the connections around the data.

Connections are key, especially when you bring together multi-disciplinary influences. Say a philosopher, a mathematician, a clinician and a data scientist walk into a bar. What happens next? They each gain a perspective they couldn’t have anticipated. “If we translate the knowledge that already exists into innovation and value, there is a huge jump in our understanding of what we observe and our capacity to predict what we will observe,” said Abdallah. More data plus more minds equals better insights.

Bring together diverse entities. “What’s nice about research consortia – which is rare in medical research – is when you’ve completed something, it’s hypervalidated,” said Lim. “It isn’t about creating something and getting everybody to adopt it; we have all the groups already there that are looking at the same problem, and they’re working on that problem together because they all want to use it.”

Dedicate the people.

“We wanted to minimize the impact [of data sharing] on people whose day job is to develop new medicines,” said McSorley. “So we created a dedicated core of staff who have been committed to making this successful, both in terms of delivering data and documents to researchers as well as getting the rest of the organization up to speed. That’s the No. 1 thing that made us successful.

“Initially, we thought it would make the most sense to have the people closest to the study do the de-identification, for example. But it became clear that it made more sense to have a dedicated group of people who specialize in this, rather than ask someone who did one study 18 months ago to try to remember what the heck to do.” Now using three experienced anonymization programmers, GSK plans to add more to ensure readiness for future requests.

Share best practices.

When organizations share what works, everyone moves a step ahead toward common goals. For example, GSK has a half-dozen SAS® macros to anonymize the patient-level data. Some of these macros redact information, some obfuscate dates, and so on. For now, the process still requires programmer judgment as to which macro to apply to the data, but the hope is for further automation in the future.

“We’ve invested a fair amount of time and money to get where we are, but if we share our de-identification macros, we allow others to benefit from what we’ve learned, so their starting point is much further along than ours was,” said McSorley.

Do something. Do it now.

It’s easy to get stuck at the start line when trying to implement a far-reaching initiative in a changing arena that lacks consensus on standards and processes. So don’t go there. Start small, see what works and go from there.

“At Project Data Sphere, we love pilots,” said Abdallah. “We pick one narrow focus, run fast on one point and try to learn as much as possible from it. It would be naive to try to implement something in a complex environment and hope to plan everything down to every step over the next four years. … We have a joke at Project Data Sphere: ‘We don’t need to know what we’re going to do; we just need to know what we’re going to do next – and learn rapidly after that.”
“Numerous scientists have pointed out the irony that right at the historical moment when we have the technologies to permit worldwide availability and distributed process of scientific data, broadening collaboration and accelerating the pace and depth of discovery . . . we are busy locking up that data and preventing the use of correspondingly advanced technologies on knowledge.”

John Wilbanks, Chief Commons Officer, Sage Bionetworks

A Timely Announcement

The greatest push for clinical trial data transparency has been from the European Medicines Agency (EMA), which in 2012 announced a commitment to complete transparency regarding patient-level clinical data and study results. On the day of the SAS Clinical Trial Data Transparency Forum, the EMA made a landmark announcement that it had decided to publish the clinical reports that underpin decision making on medicines.²

The ruling applies to clinical reports contained in all applications for centralized marketing authorizations submitted after Jan. 1, 2015. Reports will be released as soon as a decision on the application has been made. The public can either browse or search the data on screen, or download, print and save the information. In the future, EMA also plans to make individual patient data available.

According to a public statement issued by the agency that day, “EMA expects the new policy to increase trust in its regulatory work, as it will allow the general public to better understand the Agency’s decision-making. In addition, academics and researchers will be able to reassess data sets. The publication of clinical reports will also help to avoid duplication of clinical trials, foster innovation and encourage development of new medicines.”

Closing Thoughts

“The train has left the station,” said Bryson. Momentum is building. Take ClinicalStudyDataRequest.com, for example. Initiated by GSK, the online portal now represents 10 sponsors - Bayer, Boehringer Ingelheim, GSK, Lilly, Novartis, Roche, Sanofi, Takeda, UCB and ViiV Healthcare. The site receives an average of 900 visitors a day, with 219,000 unique visitors in the last 18 months, reported Dr. Jessica Scott, Director of North America Medical Advocacy and Policy at GSK.

We are moving away from “how do we think this could work” to “how is it working” and “how can we improve this as we move forward?” We are moving from questions and vision to decisive action - possibly just in time.

“We’ve overcome some of the cultural lock-in, the inertia in industry over the past couple of years since we’ve started this process – and we have gone from commitments to implementing a system that’s actually up and running and working,” said Scott.

In fact, clinical trial data transparency has moved from concept to reality quite rapidly. We’re beyond infancy in some areas, now evolving from small pilots to a sustainable model, and applying early lessons to scale, respond to feedback, and accommodate the needs of the broader industry and mutually shared goals.

“When people sign up to be subjects in an experiment, they make a tremendous sacrifice on society’s behalf,” said Krall. “Our responsibility is to make sure that sacrifice gets the best possible use. If the data can make a contribution – even if it’s a use not envisioned until years later – we have an obligation to make that possible.”

Freedland channeled his inner JFK to further remind us of the bigger picture: “Ask not what are the risks of data sharing, but what are the risks of not sharing data.”

“Part of the mindset change is that it’s not our data; it’s data that belongs to the larger medical community.”

Paul McSorley, Director, Clinical Statistics, GlaxoSmithKline

² European Medicines Agency press release, Publication of clinical reports - EMA adopts landmark policy to take effect on 1 January 2015 - Oct. 2, 2014
Learn More

Find out more about SAS and clinical trial data transparency: sas.com/lifesciences.


Read related blog posts at sas.com/hlsblog.