

## **Current and Ongoing Data Transparency Activities in the Pharmaceutical Industry**

Brief of the Multi-Regional Clinical Trials Center at Harvard University to the  
Standing Senate Committee on Social Affairs, Science and Technology

October 1, 2014

The movement toward increased transparency of clinical trials data has accelerated during the past decade as regulators, pharmaceutical companies, and journal editors have undertaken efforts to permit data generated in clinical research to be shared with other researchers or made available to the public at large. These efforts have expanded over time from the provision of *summary level* or *pooled results* to the making available of *participant-level* datasets. Below we provide a brief history of the move toward data transparency, summarize current data transparency initiatives, and offer thoughts on the risks and benefits of increased data transparency.

### **I. Data Transparency History**

The current clinical trials data transparency initiatives began in the late 1990s with the introduction of clinical trial registration requirements. The first such requirement appeared in 1997 when the United States (“U.S.”) Congress enacted the Food and Drug Administration Modernization Act (“FDAMA”), which mandated the establishment of a database of clinical trials of drugs treating life threatening conditions.<sup>1</sup> The purpose of the database was to provide a resource for patients suffering from serious diseases to find a clinical trial of an experimental treatment for their disease in which they might enroll. Accordingly, the database contained eligibility criteria for participation in the trial, a description of the location of trial sites, and the point of contact for the trial. FDAMA’s statutory mandate regarding creation of a database of clinical trials was fulfilled in 2000 through the creation of the ClinicalTrials.gov registry of clinical trials.

Clinical trials registration requirements expanded throughout the 2000s as additional countries began to require trial registration. In 2004 the World Health Organization (“WHO”) established the International Clinical Trials Registry Platform (“ICTRP”) to serve as a network of national clinical trials registries and provide international standards for such registries. In

---

<sup>1</sup> See Pub. L. No. 105-115, § 113.

connection with the creation of the ICTRP, WHO declared that “[t]he registration of all interventional trials is considered to be a scientific, ethical and moral responsibility.”<sup>2</sup> One year after the establishment of ICTRP, the International Committee of Medical Journal Editors (“ICMJE”) adopted a new policy requiring clinical trial registration in a publicly-available database as a condition of later publication of study results in any of the ICMJE-affiliated journals. Because the ICJME holds great sway over the activities of biomedical researchers due to the need to publish results in journals adhering to ICJME standards, its policy added force to the clinical trials registration movement and laid the groundwork for later data transparency initiatives. However, at that point, the ICJME’s policy did not impose *data sharing* requirements.

## **II. Data Sharing Initiatives**

### **a. Governmental Initiatives**

One of the first governmental initiatives directly aimed at data transparency was the U.S. National Institutes of Health (“NIH”) 2003 *Statement on Sharing Research Data*. The NIH policy required that NIH-grant applicants seeking a budget in excess of \$500,000 in any single year include a detailed *data-sharing plan* with their application, or in the alternative explain why such data sharing was not possible. NIH indicated that the purpose of the policy was to support the “expedited translation of research results into knowledge, products, and procedures to improve human health.”<sup>3</sup> Importantly, NIH’s policy recognized that sharing of data from research involving human participants could lead to privacy concerns, and therefore required that the rights and privacy of participants in NIH-sponsored research be protected, including through requiring removal of identifiers that might otherwise permit re-identification of individual participants.

In 2007, the United States took additional steps toward data transparency with the enactment of the Food and Drug Administration Amendments Act (“FDAAA”) of 2007. FDAAA expanded the clinical trials registration requirement of FDAMA to nearly all drug and device clinical trials.<sup>4</sup> Most importantly from the standpoint of transparency, however, FDAAA introduced for the first time a requirement that *summary results* of registered trials that form the primary basis of an efficacy claim or are conducted after the drug or device tested has received

---

<sup>2</sup> See WHO, Trial Registration, [http://www.who.int/ictrp/trial\\_reg/en/](http://www.who.int/ictrp/trial_reg/en/) (last visited Sept. 28, 2014).

<sup>3</sup> See NIH, *Final NIH Statement on Sharing Research Data*, Notice Number: NOT-OD-03-032 (Feb. 26, 2003).

<sup>4</sup> See FDAAA, Pub. L. No. 110-85, § 801 (2007), 42 U.S.C. § 282(j).

U.S. Food & Drug Administration (“FDA”) clearance or approval be posted on ClinicalTrials.gov.

In the summer of 2013, governmental policies related to data transparency entered new territory with the first proposals to mandate the sharing of *participant-level* as opposed to merely *summary* data from clinical trials. In June 2013, the European Medicines Agency (“EMA”), the European Union’s drug regulator, released a draft policy requiring data sharing for clinical trials that are used to support a marketing authorization for a drug with the EMA.<sup>5</sup> The draft policy provided that data without what EMA termed “protection of personal data” (“PPD”) concerns would be made freely available to the public.<sup>6</sup> Data possessing PPD concerns would be available upon researcher request, and requesters would be required to enter a data-sharing agreement binding the researcher to several conditions, including a pledge not to attempt to re-identify study subjects and to limit their use of the data to specific, declared research purposes. Importantly, the EMA policy did not indicate who would be responsible for enforcing the terms of such data-sharing agreements and did not provide for sanctions for failure to abide by the terms of the agreements. The EMA policy also provided that “commercially confidential information” (“CCI”) would be exempted from release.

After receiving over 1,000 comments on its draft policy during the summer and autumn of 2013, EMA has issued conflicting signals on the precise contours of its final policy. While it has not publicly released an updated draft policy, reports emerged in May 2014 that EMA was considering a policy whereby researchers would receive only “on-screen” access to data via an interface provided by the EMA, a restriction which would decrease the utility of the data to researchers by preventing downloading and manipulation of the data.<sup>7</sup> More recently, however, EMA has indicated that it will permit researchers to download, save, and print trial data for research purposes.<sup>8</sup> A final policy is expected in October 2014.

---

<sup>5</sup> See EMA, *Publication and Access to Clinical-Trial Data: Draft for Public Consultation*, Policy/0070 (June 2013).

<sup>6</sup> The EMA defines “personal data” as any data related to an identified or identifiable natural person, meaning a person who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his or her physical, physiological, mental, economic, cultural, or social identity.

<sup>7</sup> See Letter to EMA Director, Guido Rasi, EMA Policy on Publication of and Access to Clinical-Trial Data (May 13, 2014), available at <http://www.ombudsman.europa.eu/en/resources/otherdocument.faces/en/54347/html.bookmark>.

<sup>8</sup> See European Medicines Agency, European Medicines Agency agrees policy on publication of clinical trial data with more user-friendly amendments (June 12, 2014), available at [http://www.ema.europa.eu/ema/index.jsp?curl=pages/news\\_and\\_events/news/2014/06/news\\_detail\\_002124.jsp&mid=WC0b01ac058004d5c1](http://www.ema.europa.eu/ema/index.jsp?curl=pages/news_and_events/news/2014/06/news_detail_002124.jsp&mid=WC0b01ac058004d5c1).

Separate and apart from the EMA's draft data sharing policy, in 2014 the European Union ("EU") took another step toward greater transparency with the passage of a new Clinical Trials Regulation, which will be binding on all EU member states when it takes effect in 2016.<sup>9</sup> The preamble to the Clinical Trials Regulation indicates that increasing clinical trials data transparency is one of its primary goals. The regulation attempts to meet this end through the creation of a database to house (1) all information submitted as part of the clinical trial application process, (2) *summary results* for all clinical trials conducted in the EU, regardless of outcome, and (3) clinical study reports for all clinical trials used to support a marketing authorization within 30 days of the decision on the marketing authorization. Importantly, the clinical study reports contain participant-level data in their appendices.

Around the same time as the release of the EMA's draft data sharing policy, the U.S. FDA also issued a request for comments on a proposal to make *participant-level* data from medical product applications publicly available.<sup>10</sup> FDA's proposal indicated that any data released would be both "de-identified" and "masked" (*i.e.*, de-linked from a particular product). While "masking" prevents such data from being used to assess further the efficacy of a particular product, FDA's proposal notes that such data can be used to advance public health through building disease progression models, characterizing natural history of illnesses, assessing population responses to standard-of-care products, or characterizing risk factors. At the present time it remains unclear whether FDA will move forward with this proposal.

Most recently, in August 2014 the U.S. NIH announced an increase in its data sharing requirements with the release of a *Genomic Data Sharing Policy* that requires a data-sharing plan for all NIH-funded research generating large-scale human or non-human genomic datasets.<sup>11</sup> Such datasets are to be contributed in de-identified format to an NIH-designated data repository. Depending on the terms of the informed consent under which the data were collected, the data may be accessible to secondary researchers via a controlled-access mechanism, requiring review of a research proposal by a data access committee, or data may be made freely available for download.

---

<sup>9</sup> See Regulation (EU) No. 536/2014 of the European Parliament and of the Council on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC, 2014 O.J. (L 158) 1, 27.5.2014, available at <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014R0536&from=EN>.

<sup>10</sup> *Availability of Anonymized Non-Summary Safety and Efficacy Data; Request for Comments* 78 Fed. Reg. 33,421 (June 4, 2013).

<sup>11</sup> See NIH Genomic Data Sharing Policy, Notice Number: NOT-OD-14-124 (Aug. 27, 2014).

### **III. Efforts by Industry and Other Private Organizations**

In addition to governmental policies concerning data transparency, there have been a variety of private efforts undertaken in the past five years to increase data transparency.

#### **a. Company-Specific Platforms**

In 2004, pharmaceutical company GlaxoSmithKline (“GSK”) introduced a publicly accessible clinical trial registry containing summaries of GSK’s clinical research studies, becoming the first pharmaceutical company to undertake such a step.<sup>12</sup> In late 2012, GSK announced that it would make anonymized *participant-level* data available to researchers under a controlled-access model. GSK introduced several safeguards to ensure that any data released under the policy would be used in a scientifically rigorous manner. Accordingly, when applying for data access, researchers must: (1) submit a study protocol; (2) certify the presence of a statistician on the research team; and (3) commit to publish research results. Decisions on data applications are made by an independent review panel of experts appointed and compensated by GSK. By the summer of 2014, GSK’s platform had developed into a multi-company initiative hosting data from Bayer, Boehringer Ingelheim, GSK, Lilly, Novartis, Roche, Sanofi, Takeda, UCB and Viiv Healthcare.<sup>13</sup>

#### **b. Disease-Specific Platforms**

Researchers specializing in certain disease areas have undertaken data sharing efforts of their own. One example of such a platform is Project Data Sphere, which was formed by the CEO Roundtable on Cancer in early 2013 to share data from the comparator arms of clinical trials focusing on cancer treatments.<sup>14</sup> Project Data Sphere obtains data sets through voluntary submission by research sponsors. In contrast to the industry platforms described above, Project Data Sphere does not require submission of a research protocol prior to providing access to data. Rather, researchers can gain access to data after meeting basic expertise requirements and agreeing to certain terms of use.

---

<sup>12</sup> See GlaxoSmithKline, Data Transparency, <http://www.gsk.com/en-gb/our-stories/how-we-do-randd/data-transparency> (last visited Sept. 28, 2014).

<sup>13</sup> See ClinicalStudyDataRequest.com (last visited Sept. 26, 2014).

<sup>14</sup> Comparator arms consist of participants who do not receive the intervention (*i.e.*, the drug or device) that is the focus of the study. Instead, participants in these arms of the study may receive a placebo or an intervention already approved for marketing.

### **c. Academic-Private Partnership Platforms**

In addition to industry and disease-specific platforms, recent years have seen the development of data-sharing partnerships between academia and industry. One such partnership is the Yale University Open Data Access (“YODA”) Project. YODA began in 2011 when media reports suggested that researchers with financial ties to medical device manufacturer Medtronic had exaggerated clinical trial results concerning the efficacy of, and underreported adverse events related to, Medtronic’s “Infuse” spinal fusion surgery product. Medtronic engaged researchers at Yale University to conduct an independent review of datasets from the Infuse trials. Yale University announced that as part of its work with Medtronic, it would make data from the Infuse trials available to researchers through a controlled-access model, similar to the industry platforms discussed above. The YODA project has since announced a similar agreement providing for release of de-identified participant-level clinical trials data from studies sponsored by a subsidiary of pharmaceutical company Johnson & Johnson.<sup>15</sup>

### **d. Medical Journal Efforts**

Beginning in January 2013, the leading medical journal *BMJ* (formerly the *British Medical Journal*) started requiring as a condition of publication that authors commit to sharing, upon “reasonable request,” “all anonymized data on individual patients on which the analysis, results, and conclusions reported in the paper are based.” To ensure process transparency, the policy required all data requesters to make the requests public and submit re-analysis protocols to the primary authors. Authors who deny data requests are required to provide *BMJ* with a brief explanation of the reasons for the denial.<sup>16</sup>

## **IV. Benefits and Risks of Increased Data Transparency**

Increased data transparency in the pharmaceutical industry carries with it many benefits. First, it may permit the results of clinical research to be validated, as the YODA project seeks to do. This type of validation can provide health care providers and patients increased confidence that a given treatment is effective and that the list of side effects provided by the manufacturer is accurate. Secondly, the sharing of data allows for researchers to use the data to pursue secondary research that goes beyond the purpose of the project for which the data were originally collected.

---

<sup>15</sup> For more on the YODA project, see Center for Outcomes Research & Evaluation (CORE), YODA Project, <http://medicine.yale.edu/core/projects/yodap> (last visited Sept. 28, 2014).

<sup>16</sup> See Fiona Godlee, *Clinical Trial Data for All Drugs in Current Use*, 345 *BMJ* e7304 (2012); Fiona Godlee & Trish Groves, *The New BMJ Policy on Sharing Data from Drug and Device Trials*, 345 *BMJ* e7888 (2012).

Allowing data that have already been collected to be used for secondary purposes can speed up scientific discoveries by relieving investigators of the burden of collecting new data. Such secondary research may lead to promising new treatments or a better understanding of disease through various “big data” projects.

While increased data transparency offers many possible benefits, it also carries with it several risks. The move toward increased release of participant-level data carries with it the risk of re-identification of study participants. Accordingly, all of the examples of data sharing proposals discussed above that permit the release of *participant-level* data require some form of anonymization prior to release of the data. Nevertheless, given the increasing ability of genetic and other types of data to be re-identified through comparison with information in public databases, de-identification requires the removal of ever larger lists of “identifiers” in order to be accomplished successfully. Unfortunately, the removal of identifiers is inversely proportional to the utility of the data (*i.e.*, as more identifiers are removed, researchers have less ability to make meaningful use of the data).

A second risk posed by increased data transparency is that it will provide pharmaceutical companies opportunities to exploit commercially data that have been released by their rivals. The release of data may permit rival companies to use information released as a result of data transparency legislation in one jurisdiction to seek regulatory approval of competing products in other jurisdictions without incurring associated research costs. Increased transparency can also lead to competitors learning about research and commercial strategies of a given company. For example, requiring the release of data from failed products can provide competitors insight into other possible research pathways.

A third risk is that data will be used by researchers who lack the knowledge and skills to make productive and sound use of the data but who nonetheless are able to attract public attention to their data analyses. This could result in unqualified researchers using flawed analyses to attack manufacturer efficacy claims or challenge the decision of regulators regarding a given decision on a marketing authorization. Rather than advancing public health, flawed analyses can confound scientific and public knowledge of treatment options. Additionally, the proliferation of such analyses can increase the burden on overworked regulators, as they must expend time analyzing erroneous data that have been submitted to challenge earlier regulatory decisions regarding a given product.

Taken together, the risks and benefits of data transparency militate toward a policy of controlled access to data, in which researchers are required to submit to a gatekeeping committee a research plan as well as evidence of their qualifications prior to receiving access to data. This model is sometimes referred to as a Learned Intermediary Model.<sup>17</sup> As shown in the examples surveyed above, this gatekeeping committee could consist of persons appointed by governmental agencies, academic institutions or pharmaceutical companies; however, any such body should operate as a neutral third party, independent of the clinical trial sponsor. The committee members collectively should possess knowledge of the scientific research process, common statistical techniques, data privacy regulations and best practices, and ethical considerations of research. Placing control over decisions to release data in a neutral party with appropriate expertise can guard against biased decisions by industry and academic sponsors of research and thus can improve public trust in the system. Such a process also offers advantages over open access models by hindering flawed analyses and ensuring that evaluations of requests for data access are based on scientific merit and the requesting party's ability to use the data for meaningful, statistically valid research purposes.

---

<sup>17</sup> For a further description of the learned intermediary model, see Michelle M. Mello, Jeffrey K. Francer, Marc Wilenzick, Patricia Teden, Barbara E. Bierer & Mark Barnes, *Preparing for Responsible Sharing of Clinical Trial Data*, 369 NEW ENGLAND J. OF MED. 1651 (2013).