Title: Regulatory Implications of Data-mining
Author: Renee Harrell
Word Count: 2387
Regulatory Implications of Data-mining

Introduction

The Food and Drug Administration (“FDA”) has expressed interest in data mining the records collected from numerous clinical research trials with the intent of advancing medical breakthroughs and improving patient outcomes with less time and expense than traditional clinical trials. While the FDA’s intended goal is meritorious, there are numerous ethical and regulatory challenges impeding the pursuit of such a goal. This paper focuses on the regulatory and ethical considerations pertaining to data ownership and informed consent. Only after data ownership and informed consent are addressed can FDA move forward with other regulatory challenges presented.

In October 2011, the FDA released a report titled Driving Biomedical Innovation: Initiatives to Improve Products for Patients in which the Commissioner identifies major reforms to be addressed across several topics. One area identified was the “harnessing the potential of data mining and information sharing while maintaining strong privacy protections.” The report proclaims, “the ability to integrate and analyze the data housed at FDA could revolutionize the development of new patient treatments and allow us to address fundamental scientific questions about how different types of patients respond to therapy.” However, the report also noted FDA had not previously been able to mine de-identified clinical trial data due to “inability to combine data from disparate sources and the lack of computing power and tools to perform such complex analyses.”

FDA houses the largest repository of clinical data. The repository consists of de-identified data submitted to the Agency for new product approval. The FDA is considering approaches to provide access to non-FDA experts to fully utilize the potential of the data. The FDA is also evaluating “the extent and nature of public availability of de-identified and masked subject level data necessary to achieve the specific aims” consistent with Commissioner Hamburg’s Transparency Initiative. FDA proclaimed that the mining of existent data could be valuable and is considering making de-identified data available to other experts and possibly the public.

This paper discusses the regulatory implications of public data-mining of de-identified clinical trials data for Patient–Centered outcomes research. First, definitions of key terms will be addressed. Next, the regulatory framework that governs existing repositories of data housed by the FDA is addressed. Finally, this paper presents recommendations for handling current data as well as the ethical treatment of future subjects and their resultant data.

Definitions

Consistent definitions are important in understanding the regulatory implications and concerns facing FDA’s data mining proposal. This paper utilizes the following defined terms:

- Public – Relating or belonging to an entire community, state, or nation. Open or available for all to use, share, or enjoy.
- De-identified - Application of 45 C.F.R., §164.514(a) of the Privacy Rule and appropriate methods used for de-identification are those described in Section 164.514(b) of the Privacy Rule.
- Data Mining – “the use of complex data analytics to discover patterns of associations or unexpected occurrences (‘signals’) in large databases. When such signals are identified, they can then be evaluated for potential intervention as appropriate, such as further assessment (‘signal refinement’), labeling revisions, and hypothesis testing studies. Much of the data mining work at FDA involves detecting safety signals among adverse events reported to the Agency.”

Data Ownership

Before the FDA makes a decision on mining the data in their repository, it is crucial to come to a consensus on who owns the data in the repository. "Basic principles regarding ownership of data is not open to debate as a legal matter." According to Fishbein, “The regulations governing federally supported research usually permit grantee universities in research institutions to retain legal title to scientific data.” Fishbein also states that “even apart from regulations affecting government supported work, the principle that ownership of primary research data, in whatever form it may be expressed, resides in the employing institution rather than in the employee or faculty member.” Until a formal definition on data ownership is established the research community must look to the use of the term ‘ownership of data’ by the FDA in order to derive its meaning.

While FDA has no direct guidance on the subject, 21 C.F.R. § 314.50 states that “...the applicant shall include in its application a written statement signed by the owner of the data from each such investigation that the applicant may rely on in support of the approval of its application, and provide FDA access to, the underlying raw data that provide the basis for the report of the investigation submitted in its application.” (emphasis added) In this it is clear that FDA differentiates between access to and ownership of data. It can be inferred from the above use of the phrase “owner of the data’ that FDA does not consider themselves owners.

When investigating data ownership, consideration must be given to how and why the data is obtained. It is apparent from the statement above that FDA has access to but not ownership of the data. FDA obtained the data housed in their repository through mandatory submissions required and enforced by FDA. FDA obtained these data from the sponsor as a requirement for receiving study approval. FDA did not, however, obtain ownership of the data. Because FDA does not legally own the data in their repository, they must receive permission from the data owner prior to using or granting access to data that otherwise may not be considered public.

FDA must consider the legality of using the data for purposes other than is the purpose stated in current regulations or the regulations at the time the data was obtained by FDA. In regards to ownership of data, FDA does not have the legal right to use the data they currently house for data-mining.

Informed Consent

Informed consent is the hallmark of ethical clinical trial design and practice. This notion is embodied by 21 C.F.R. § 50.20 which declares “no investigator may involve a human being as a subject in research covered by these regulations unless the investigator has obtained the legally effective informed consent of the subject or the subject's legally authorized representative.” Clinical trial data can only be gathered once a trial subject gives informed consent to the researcher. The data currently housed in FDAs repository is from patients that acknowledged their informed consent to participate in a clinical trial, however, there was no indication made (with the exception of Clinicaltrials.gov studies) that any de-identified research data would be made available to an outside party, especially the Public, for the purposes of data-mining.

If FDA were to mine the data retained in their repository, they would be doing so without the informed consent of the subjects. This would be in direct violation of their own regulations specifically, the elements of informed consent stated in 21 C.F.R. § 50.25(a).

• 21 C.F.R. § 50.25(a)(1) states “A statement that the study involves research, an explanation of the purposes of the research and the expected duration of the subject’s participation, a description of the procedures to be followed, and identification of any procedures which are experimental.” Trial subjects that are the subjects of data mining will never receive a comprehensive explanation of the purpose and expected duration of their participation. Any informed consent given by the subject at the time of the study
could not have contained language indicating that the research data collected would be used for an indeterminate amount of time by FDA or be used for data-mining.

- 21 C.F.R. § 50.25(a)(2) states “A description of any reasonably foreseeable risks or discomforts to the subject.” The subject could not have been adequately informed about foreseeable risks as required because the mining of de-identified data by FDA has not yet occurred.

- 21 C.F.R. § 50.25(a)(5) stipulates that “A statement describing the extent, if any, to which confidentiality of records identifying the subject will be maintained and that notes the possibility that the Food and Drug Administration may inspect the records” be included in the informed consent. It is likely that any informed consent waiver did not accurately inform the subject about the confidentiality of records and extent to which FDA may use the records.

- There is also the possibility that the use of language required by 21 C.F.R. § 50.25(a)(5) requiring a statement that the FDA may inspect the records may be considered misleading, because it states that FDA may inspect the records which is vastly different from FDA using records for the continuation of research purposes.

FDA repeatedly touts that the data contained in their repository is de-identified to protect subjects’ privacy. De-identification of data does may make the endeavor of data mining low-risk or no-risk, but does not remove the necessity or purpose of informed consent. On January 27, 1981, FDA made comments on the final rule titled, Protection of Human Subjects; Informed Consent, 21 CFR Parts 50, 71, 171, 180, 310, 312, 314, 320, 330, 361, 430, 431, 601, 630, 812, 813, 1003, 1010 [Docket No. 78N-0400] 46 FR 8942 supporting that notion. In response to questions about exempting no and low-risk studies from informed consent, FDA responded in agreement with the position stated by The National Commission “that even in no-risk or low-risk studies, respect for the rights and dignity of human subjects would require informed consent before participation in any clinical investigation.” The FDA further commented that “The Federal Food, Drug, and Cosmetic Act requires that these regulations have due regard for the interests of patients (21 U.S.C. 355(j)(1) and 21 U.S.C. 357(g)(1)) or be consistent with ethical standards (21 U.S.C. 360(j)(g)(1)). Therefore, FDA believes it possesses the necessary statutory authority to reject studies where informed consent has not been obtained even though the scientific validity of the data generated may not have been affected, and it reserves the right to do so where circumstances so warrant.”

The FDA considers the use of data, including de-identified data, be disclosed to human subjects as part of a valid informed consent. This is evidenced by the specific language requirements contained in any informed consent for trials defined in 42 U.S.C. 282(j)(1)(A) whereby, all applicable “informed consent notices must contain the following statement to each clinical trial subject in informed consent documents and processes.” The statement is: “A description of this clinical trial will be available on http://www.ClinicalTrials.gov, as required by U.S. Law. This Web site will not include information that can identify you. At most, the Web site will include a summary of the results. You can search this Web site at any time.”

This statement notifies the clinical trial subject that clinical trial information has been or will be submitted for inclusion in the clinical trial registry databank under paragraph (j) of section 402 of the Public Health Service Act.

Recommendations

Current data housed at FDA should be handled in accordance with current regulations. Only data obtained with the proper consent should be used for data mining. These data consist of data obtained from trials defined in 42 U.S.C. 282(j)(1)(A) that require reporting on clinicaltrials.gov. This is the only data available to the FDA and the public that could legally and ethically be used for mining. However, the issue with using this data is many trials required to report their results do not in fact submit their data. One study examined compliance with
mandatory reporting of clinical trial results on clinicaltrials.gov and concluded that only 22% of the studies mandated to report results on clinicaltrials.gov did so. This means the FDA currently has a database whereby results from clinical trials are mandated to be reported for data analysis but FDA is not enforcing its existing regulations. Therefore, my first recommendation is to enforce the current rule regarding data submittal to clinicaltrials.gov and begin mining that data. The FDA has proven they cannot consistently collect data and maintain a database that requires cooperation from the research community because they do not or cannot enforce compliance. Devoting already limited resources on building a new system may be ill advised.

At this time, FDA should focus their efforts on ensuring that all data required to be submitted to FDA is submitted. FDA should enforce this existing rule and use this properly obtained data as a platform to test their data mining program, prior to granting itself authority to allow for expanded data-mining. The FDA should not be collecting data for research purposes as such activities are beyond their scope of mission. The idea of data mining has value but should more properly fall under the auspices of National Institutes of Health ("NIH") as such activities are more congruent with the purpose and mission of NIH than FDA. NIH could offer this feature as an expansion of the clinicaltrials.gov website. The expanded portion of this database would be made available to persons only after an official request is made to NIH. The official request for information for research purposes would need to state the intended purpose of the research, as well as the results from their data analysis once completed. A designated committee within NIH should be responsible for approving or denying such request and FDA would be responsible for enforcing regulations related to making the data findings available.

The initial step to mining current de-identified clinical trial data is to use data from studies where subjects receive informed consent that includes the extent to which the data would be disclosed to the FDA and public, including instances of data mining. This data mining could retrieve data obtained from trials operating under 42 U.S.C. 282(j)(1)(A). As stated above, FDA needs to enforce the mandatory reporting requirements for these trials in effort to both increase the data pool, as well as to test FDA’s ability to enforce compliance with data reporting, before they extend their efforts to a larger group. The second step is to change the requirements of informed consent. Changes should require language similar to that in 42 U.S.C. 282(j)(1)(A) be inserted into all consent forms informing the subject about the use of their data by FDA. No data, from any clinical trial, should be used for data mining or made public unless the study participant is made aware of the intentions as well as the risks. Finally, the issue of data ownership would become moot if all parties are aware and the subject grants proper informed consent to the researcher thereby avoiding the necessity to obtain permission for data mining use by the data owner.
References


ii Id. at 24.

iii Id. at 22.


v Id.


ix Id.

x Id. at 129.

xi 21 C.F.R. § 314.50

xii 21 C.F.R. § 50.20

xiii 21 C.F.R. § 50.25(a)(1)

xiv Id.

xv 21 C.F.R. § 50.25(a)(5)


xvii Id.

xviii 42 U.S.C. § 282(j)(1)(A)