Future Developments in Clinical Studies: Big Data Analysis

Olivier Goarnisson

The Potential Benefits of Big Data Analysis

Big Data refers to the analysis of massive, quasi-unlimited amounts of heterogeneous data. There is currently a great deal of excitement in the life sciences industry about how clinical researchers might use Big Data techniques to analyze data from much larger patient groups than those that can be included in clinical trials. Using Big Data in this way could potentially be game-changing in reducing bias in results. This is because the group of patients that can be included in a clinical trial is only a sample of the actual group of patients whom the treatment that is being trialed could benefit. But the selection of a small group of patients on whom to conduct a clinical trial can give rise to bias in a trial’s results if patients are selected in such a way that the investigational treatment in question is more likely to be successful. It is also possible to select a patient group in a way that generates fewer adverse events to the treatment than would be the case under real world conditions, for example by including a higher proportion of young, fit patients than exists in the general population. Big Data analysis can to a certain extent help researchers reduce these types of clinical trial biases. It can also help to reduce or even exclude some of the ethical problems that are currently inherent in the conduct of clinical trials.

The Current Problems with Big Data Analysis

Researchers are not yet able to effectively analyze the massive amounts of heterogeneous data that exist in a Big Data pool. The technologies and scientific methodologies that are currently available to the life sciences industry do not yet allow researchers to extract the data relevant to their studies from such a massive data set, and then to properly structure these data so that they can be analyzed in a meaningful way. As a result, the research projects that have taken place so far have been too limited in scope to be considered a comprehensive Big Data analysis. However, some have come close to fully fledged Big Data analysis by attempting to pool data bases from multiple clinical research projects and making these data available to scientists who wish to access them as part of their research. The use of such data pooled from many clinical trial databases is already helping with the conception of new clinical trials.

Examples of Attempts at Using Big Data Analysis to Conceive of New Clinical Trials

The initial attempts of researchers to come close to Big Data analysis have already made some progress towards addressing two problems that were previously believed to be inherent to clinical trials. The first problem is the relatively long time that one has to wait for...
the results of the clinical trial of, for example, a cancer treatment. The second involves the ethical issues that can arise during clinical trials of treatments for rare diseases.

The efficacy of a treatment for malignant tumors is usually measured on the basis of how many of the patients included in the clinical trial have survived five years after the trial has concluded. Having to wait for five years before having enough overall survival data to be able to conclude that the treatment under investigation is effective means a longer wait for patients before a cancer medicine becomes widely available on the market. It also means a longer delay before a pharmaceutical company can begin to earn money from the new drug.

However, in a clinical trial, data about how many patients have survived free from any further progression of their cancer are available for analysis well before the five-year mark. By attempting to use Big Data analysis techniques on the results of a large number of clinical trials that have already been completed, researchers have succeeded in demonstrating to the satisfaction of the scientific community (though not yet to that of the regulatory authorities) that, for certain malignant tumors, progression-free survival rates at the one-year mark give a good indication of what overall survival rates are likely to be after five years. Based on these researchers’ analyses, it is possible in clinical trials of treatments for malignant tumors to reach a preliminary conclusion about the efficacy of the treatment under investigation using the data available after only one year.

There have already been attempts at Big Data analyses that have managed to address some of the ethical problems arising in clinical trials of rare diseases. A traditional clinical trial is conducted on two groups of patients, one group that takes the drug that is being investigated (arm A) and one group that takes another medicine or a placebo (arm B) for comparison. However, in the case of certain rare diseases, the overall patient population is so small that one cannot have two groups of patients. An additional consideration is that, when investigating treatments for serious and fast-progressing illnesses, it is unethical to put an arm B group on a placebo.

However, there have recently been clinical trials in which the regulatory authorities have allowed arm B to be a group that has been virtually generated by using Big Data analysis techniques on pooled data from electronic health records, registries and earlier clinical trials. This data analysis has then been used to set up the arm B of a trial. In these cases, the relevant data were identified in accordance with the research protocol and then extracted from earlier clinical trial databases, registries and electronic healthcare records and transferred to the study database of the new clinical trial to compile a virtual arm B patient group for comparison with the arm A group of actual patients.

The projects in these examples have been restricted in scope, and the regulatory authorities have not yet fully approved their methods as a valid support for a marketing authorization application. They do represent real progress in the field however. In the future, it is likely that more comprehensive analyses of Big Data sets will be used to assist in the conduct of a significant number of clinical trials. Some market commentators have even suggested that
Big Data analysis will ultimately fully replace the classical clinical trial. Although this is very unlikely to happen, particularly when the trial in question is of a new compound. A drug that is new to the market will, of course, never have been tested. There will therefore be no data relating to it in the Big Data pool, so a purely Big Data-based trial would be impossible in such a situation.

Concerns About the Use of Big Data

As the pool of Big Data from the results of clinical trials grows, some people fear the development of a scenario in which the health data of an unlimited number of people could be used to support clinical trials or develop medicines. However, in practice, both the sponsors of clinical research who wish to use data from health records and the institutions who agree to provide them with such data are subject to very strict data protection laws, especially within the European Union.

A Big Data set for the life sciences industry must be grown responsibly, in a way that allows it to be useable from both an ethical and regulatory standpoint. For this reason, it is important that clinical trials generating results that may end up in the pool of Big Data have proper support from a multidisciplinary legal team that is familiar with clinical trial regulations, intellectual property law and data protection law.