MDIC Patient-Centered Clinical Trial Design Workshop – May 18, 2018

Aim 3 Breakout Session: Designing Methods for Clinical Trials Based Explicitly on Patient Input

Shomesh E. Chaudhuri led a breakout session that included an overview of the methods used by the MIT Laboratory for Financial Engineering to incorporate patient preferences into statistical significance thresholds in clinical trial design. Chaudhuri explained how the MIT team used a patient value model that provides the foundation for a Bayesian Decision Analysis ("BDA") and illustrated the framework's application in a case study for a neurostimulator device. A panel discussion on these novel methods followed, featuring Martin Ho (FDA), Kevin Kwok (MJFF), and Edward Karst (Abbott).

Patient Value Model

The patient value model uses patient preferences to estimate the value associated with three considerations in clinical trials: type I errors (false approvals of products that are ineffective and/or harmful), type II errors (false rejections of products that are safe and effective), and trial length. Type I errors were accounted for in the model by using three risk factors identified by Parkinson's patients: additional risk of depression, risk of brain bleeds, and mortality risks. Patient preference information regarding these three risks was used to estimate the relative importance of these risk factors for decision-making. This risk profile allowed for the estimation of the average value lost for patients who were using an inferior treatment compared to the standard of care.

The patient value model also considered value lost from missed opportunities to receive a therapy that is more effective than the standard of care (type II errors). Patient preferences were used regarding five criteria: (1) whether the patient received deep brain stimulation ("DBS") treatment previously, (2) whether the patient has experienced dyskinesia, (3) severity of motor symptoms, (4) ambulation, and (5) severity of cognitive symptoms.
Clinical trial length was also factored into the model through discounting, allowing the model to consider the benefit to patients from receiving a treatment sooner. The discount rate was found to increase with age and severity of symptoms, meaning that older patients with more severe symptoms placed higher value on receiving treatment sooner compared to younger patients with less severe symptoms.

Bayesian Decision Analysis & Application

The patient preference information allowed for the estimation of the value lost in both false approval and false rejection scenarios while taking into account the number of patients required for clinical trials. Estimations of the probability of each scenario were then calculated using statistical power (derived from the size of the sample population) and statistical thresholds. The BDA then used these probabilities to maximize the expected patient value for various types of patients.

The MIT team applied this framework to the design of a clinical trial for a hypothetical neurostimulation device. A numerical benefit-risk profile for the therapy was generated using values from the literature, and the BDA was run for various subgroups of patients with varying risk tolerances. Patients who had more severe symptoms and who had already experienced DBS were generally more risk tolerant. The BDA-optimal situation for these patients was to allow for a type I error rate closer to nine percent (rather than the traditional five percent), and to allow for shorter trials to expedite approval. The opposite was true for patients with less severe symptoms who had not experienced DBS. Those patients were less risk tolerant, and the BDA-optimal situation was to restrict type I error rate below one percent.

In this case study, the traditional fixed significance level of five percent did not maximize patient value for most of the patients. For patients with more severe symptoms, value was lost in lengthy clinical trials that were overprotective of false approvals. On the other hand, more risk-averse patients with less severe symptoms would have benefited from a threshold with a smaller p-value; the traditional threshold of five percent was too permissive given the risk profile of the device.

Panel Discussion

Kevin Kwok (MJFF) started the panel discussion by sharing a story about his Parkinson's diagnosis and the choices he made regarding his election to get DBS. Kwok explained that his Parkinson's symptoms were not severe, but that he elected to get DBS earlier than was recommended by his physicians, and against the wishes of his family.

Edward Karst of Abbott underscored the difficulty of convincing clinicians and others to abandon the traditional fixed p-value threshold of 0.05. He also pointed out that manufacturers may not be willing to conduct the up-front research on patient preferences that informs the patient value model. He also emphasized that alternative p-value thresholds must be established before the relevant clinical trial is conducted, not after a trial is conducted to justify a p-value that would be deemed insignificant because it is greater than 0.05. Kwok highlighted the need to educate and reward companies that use new types of trials, rather than have them use the same types of trials because they know that's the best way to get approval.

Andrew Lo (MIT) echoed Karst's concerns about the barriers of cultural change with regards to the default 0.05 p-value threshold. He also referenced a comment that Kwok made in a discussion from earlier in the day regarding the emotional burden of recommending a treatment to a group of people, some of whom would die from that treatment. Kwok's initial remark framed that treatment recommendation as a bad decision because patients died, but Lo emphasized that people are going to die whether they receive a treatment or not, and that the most important thing to focus on is to ensure that those deaths are not for nothing. He brought up the fact the emotional burden of these decisions is also a consideration for FDA reviewers. He argued that putting together an objective, systematic, and repeatable framework where all stakeholders decide together what the appropriate values are for approving or not approving therapies is a more ethical approach than the status quo.

Murray Sheldon (FDA) agreed, stating that the emotional burden creates a moral dilemma that drives FDA reviewers to be risk averse. However, he stressed that deciding to not approve a product can still result in people getting hurt, and that risk may be equal to or worse than the opposite approval decision, even though it does not confer an emotional burden on reviewers.

Martin Ho (FDA) pointed out that this type of model is flexible enough to be used on other thresholds beyond the p-value. Other audience members were interested in the flexibility of the framework to guide other binary decisions, such as whether a patient should choose to use a therapy, and simultaneously incorporating preferences from multiple stakeholders, for example, parents and children.

Karst emphasized the efficiency of being able to test medical devices in a single investigation, instead of using the traditional multi-stage process. He also brought up the possibility of partial approval based on severity of symptoms, rather than having an all-or-nothing approval decision for all patients.

Andrew Lo suggested that a partial approval decision may work similarly to the system used in certain types of investing. He explained that regular investors are normally prohibited from taking on certain types of investments, but individuals who are knowledgeable about those types of investments and who have a certain degree of net worth are given an exception. Similarly, Lo suggested that patients who demonstrate that they have the appropriate knowledge to assess the additional risk of an investigational treatment could be allowed to participate in these trials.