CIS 2 Alaleh Azhir
Seminar Critical Review 3/10/18 Team 5

Introduction:

Physicians lack a simple method to quantitatively analyze and perform exploratory analysis of dose distributions within organs before, during and after treatment. The goal of our CIS 2 project is to create a user interface with 3D rendering that allows physicians to better analyze dose distributions before treatment and make necessary adjustments to the treatment plan. We are currently developing a web-based user interface, that 1) utilizes a python built backend that communicates with a SQL databased to obtain medically relevant data 2) creates interactive 3d visualizations of organs using JavaScript libraries such as D3.js, 3) allows physicians to segment organs into regions, select and drag regions, and run dose volume histogram analysis on these new regions, 4) display Dose-Volume Histograms (DVH) curved using interactive zoomable graphs, and finally allows for integration of new feature analysis scripts to be added to the interface. As a background paper, I chose "How Will Big Data Impact Clinical Decision Making and Precision Medicine in Radiation Therapy?" by Chen et al. to study the relevance and motivation behind the creation of our web-based user interface. As this paper is written by our mentor Dr. McNutt, I thought it would provide a valuable background on why our project is important.

Paper: How Will Big Data Impact Clinical Decision Making and Precision Medicine in Radiation Therapy? [1]

Background information:

This paper first explains precision medicine, and its effects on our current healthcare system. Precision medicine refers to tailoring treatment to each individual patient based on his/her genes, medical history and demographics, rather than generalizing it and only classifying it by the organ affected. Precision medicine is increasingly applied in cancer treatment, as it has transformed from a fixed therapy for all patients solely based on the organ, to a more personalized treatment based on one's genetic and molecular factors, leading to better outcomes. However, personalized treatment has resulted in a major issue: the decrease in the number of "like patients" getting the same exact treatment thus impeding our ability to test research hypotheses with sufficient statistical power. As precision medicine has become growingly adopted, at the same time Electronic Health Records have also been getting increasingly adopted (an example of "Big Data") throughout the United States. Electronic Health Records provide vast data on various patients whose treatments could be similar, as a result of simple check-ups. In the next section, the authors describe the current use of big data in clinical setting.

Big Data Use in Clinical Setting

Big data has 2 major uses in current medical practice these days:

1) One of the main uses of big data is to assess the quality of healthcare, and current clinical practices. The authors give a few examples of such studies, the highlights of which are described below:

- The studies have shown an underuse of aggressive treatment for high-risk prostate cancer patients.
- The studies have also shown an overuse of bone CT scans, and further follow ups with MRIs for patients with early prostate cancers, which has led to an expense of $12 million annually for Medicare. Although this may not seem like a significant cost, given the large population of prostate cancer patients, this is only the cost for bone scans, and does not include the other types of extraneous unnecessary treatments that early stage patients may go through.

2) Another use for big data in healthcare is to perform comparative treatments research. Despite the fact that traditionally randomized clinical are used to compare different treatments and their effectiveness on patient outcomes, big data can provide the following advantages over clinical trials:

Table 1 Advantages and disadvantages of observational cohort and clinical trial data for comparative effectiveness research

Challenges	Observational cohort ("big data")	Clinical trial
Confounding	Disadvantage	Advantage
Detailed data elements	Disadvantage	Advantage
Sample size	Advantage	Disadvantage
Timely results	Advantage	Disadvantage
Generalizability	Advantage	Disadvantage

Adopted from Chen et al. 2015 [1]

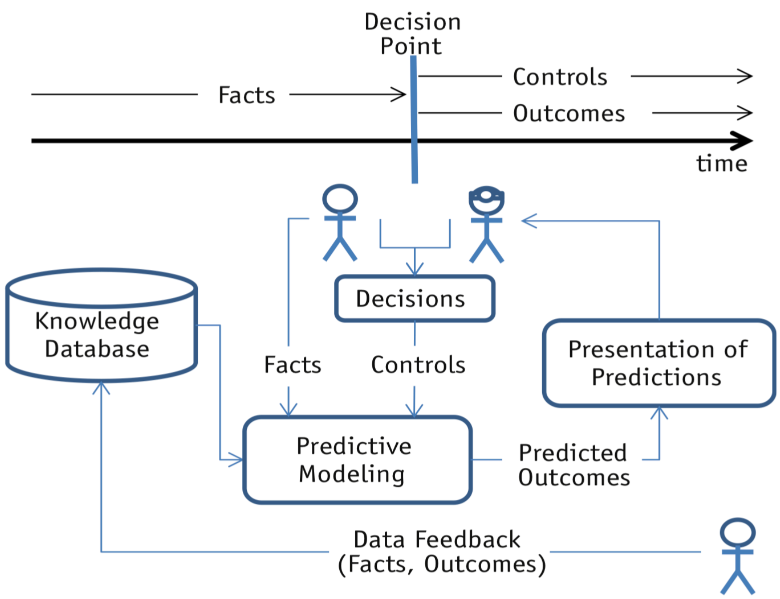
- It is more likely for a clinical trial to be affected by small sample size thus big data analysis is more likely to be generalizable, it takes much shorter amount of time to complete a big data analysis compared to clinical trials, and finally although big data can be affected by bias due to confounding variables, clinical trials control for confounding variables.

Limitations of Big Data

The authors offer an assessment of the limitations of big data which include the fact that currently, the big data available is 2-3 years outdated and before the current clinical practice. This is due to ever changing nature of clinical data as new methods of treatments for cancer develop every day. Also, for the data to go into a database, it needs to be checked for errors, normalized and standardized, which itself takes time.

Vision for Future

It has been shown that predicting achievable dose based on the organ at risk, and a knowledge database of previous treatments and the patient outcomes can improve treatment planning. Furthermore, this data should be retrieved and be open to exploratory analysis at the time of consultation to be relevant in a clinical setting, and each patient's data should be easily incorporated into the database. There are many stages of decision making in clinical practice from initial assessment of deciding the course of treatments to reevaluating the treatment given patient's reaction and symptoms. These all can be helped if the physician could use a knowledge database that contains patients' courses of treatments, genetic makeup, demographics and outcomes and a machine learning algorithm that could find the closest and best possible treatment for a current patient. Then using the predicted outcome of a certain treatment the doctor can judge what the best course of action is at each decision point. This was very nicely demonstrated in their paper using the following graph:



Adopted from Chen et. al. 2015 [1]

My Evaluation of Pros and Cons of the Paper:

Pros:

- Their vision for future provided our group with some ideas outlined below on what sort of analysis would be useful to integrate into our user interface.
- The paper provided a very good overview of what is the overall goal of our project, and how it can contribute to the growth of precision medicine.
- Through examples in clinical settings, this paper clearly showed the current beneficial uses of big data and its limitations.

Cons:

- The authors did not implement any of their vision for future in the paper or give any examples of steps being taken towards that end as of right now, but rather gave examples of currently existing methods.
- This paper does not provide a concrete example of a currently existing user interface, for us to build further on, but rather provides an overview of what can be relevant clinically. However, providing examples of user interfaces was not the aim of their paper.

Our Next Steps:

Their assessment of how big the data will be led me to research more about how to visualize such large datasets on a web-based platforms, and whether we needed to make any adjustments to our plan of action:

Big Data Visualization on the Web:

- Currently we use D3.js, which accesses Document Object Model elements (i.e. if we have 5000 points on the graph, it creates 5000 objects onto the website). Therefore, as one zooms, drags and interacts with the page, we would need to update 5000 objects immediately which can be quite slow.
- We are consider switching to Pixi.js, which is JavaScript library using WebGL and draws all elements onto a canvas (1 element). Thus, as one interacts with the object graph, it only needs to update 1 element on the page. The downside to using Pixi however is that it is more pixelated and not as nicely developed as d3.js.
- Another problem to consider may be continuous zooming, in which we can look at different tiling methods, and aggregation of our data, such as the one implemented by Google Maps.

Big Data Analysis on the Web:

- Include a function that allows for comparative data analysis across patients and organs.
- Currently we process data by passing it to the backend Python layer, as our data size grows, this analysis could take longer, thus we may want to consider using WebWorkers to stop page from crashing.

- Furthermore, we can enable users to make accounts to save progress of their current analysis, to let them exit the page and come back to it when the analysis is finished, through the use of a middleman such as Gearman or PHP.
- Also, we can instead use Electron.js to transform the code into a desktop application in the case where running the analysis takes too long.

Big data storage:

- As the data grows larger, we may want to consider various compression techniques, to save storage space, and be more efficient.

References:

[1] Chen R, Gabriel P, Kavanagh B, McNutt T, "How will big data impact clinical decision making and precision medicine in radiation therapy?" Int'l J. of Radiation Oncology, Biology, Physics. Published online: November 27 2015