Understanding Clinical Care and Clinical Data

CSCI 8980 scribed lecture

Scribe: Andrius Adomavicius

1 Introduction

This lecture introduces the general and specific tasks and goals of healthcare, as well as illustrative examples of clinical data and how to use and process it to observe and understand its meaning.

2 Goals of Clinical Care

Generally speaking, the overall goal of healthcare should be to support the health of the people. We understand **health** as an all-encompassing term referring to a state of complete physical, mental, and social well-being, rather than just an absence of illness or morbidity. Efforts in healthcare are not limited to treatment and cure of consequential diseases, though many important and high-stakes efforts are focused on this. For instance, one can consider efforts to reduce mortality, disability, and morbidity.

Across the world, advances in healthcare technology have allowed life expectancy to rise over the past generations, especially rapidly in developing countries, reducing many preventable early mortalities. However, there still remain many causes of death that have remained difficult to prevent, even among the most developed countries. For instance, heart disease and cancer have accounted for over 45% of deaths in the United States in 2014. Other chronic conditions such as arthritis and hypertension are also prevalent, and despite not directly causing as many deaths they may cause other complications and weakness and it would be in our best interest to have better treatments for these morbidities.

Time scales We may also classify different health conditions and healthcare tasks by appropriate time scales. One can have short-term acute illnesses for which a cure is sought, or long-term chronic illnesses for which efforts are made to manage and treat the severity of the illness. More proactive approaches towards disease such as prevention and prediction may also be done on patients who are not (yet) known to suffer from disease.

Treatment tasks in healthcare When dealing with diseases and impairments, we consider three key tasks in the clinical process. **Diagnosis** is the identification of a disease or condition from the signs and symptoms that have been observed and interpreted. This step is important in determining a specific disease a patient may be suffering, which will help with treatment based on knowledge and studies of the disease itself. A related task is **prognosis** which aims to predict the course or trajectory of the disease and a patient's potential recovery from it, based on the diagnosis and the other evidence observed. Given a clear diagnosis and prognosis an appropriate **therapy** can be done in the form of treatment and management of the disease, for instance in

prescribing a suitable medication given the data observed and inferences made. These processes form a medical cycle in which a patient can be further examined after the effects of the therapy, with the diagnosis and prognosis able to be updated with more confidence if needed, given the additional data observed.

This cycle can be seen through the lens of the scientific method, in which the diagnoses and prognoses are hypotheses that can be tested through therapy, which may bring more evidence in support of or against the hypothesis, allowing it to be updated appropriately.

Preventive tasks In addition to treatment for ill patients, medical knowledge should also be applied to prevent healthy people from becoming ill. This is often done for rapidly spreading epidemics, where contact tracing is a measure used to understand the spread of disease across space and time, allowing a means of assessing the risk of infection and recommendations of isolation and quarantine as preventive measures. In general disease prevalence can be better tracked across the world as a result of the standardized ICD-10 disease codes under the control of the World Health Organization, allowing for uniformity across different sources of data.

Other health system tasks In addition to such direct tasks in identifying, treating, and preventing disease, other meta-level processes across a clinic or health system are critical to overall health outcomes. For instance, there must be a robust system for acquiring and applying medical knowledge and ensuring the health workers are appropriately educated in this knowledge. There also must be strong quality control and process improvement, including management of costs. An organized archive is also important for the appropriate data and knowledge to be conveniently accessible. Additional enterprise-level tasks such as managing the membership and maintaining electronic systems are also important for a clinic, so the overall health system has many important and interdependent parts. In particular, the field of Learning Health System Sciences deals with this process as a whole, generating and applying new knowledge based on scientific, technical, and cultural changes and improvements.

Improvements Identified areas for improvement include a comprehensive integration of clinical data with social, behavioral, and demographic data, as well as capture of novel data sources (such as genomes) and improvements of technical infrastructure to store and access this data. Frequently used clinical procedures have improvements to be desired in their effects, for instance in Randomized Clinical Trials (RCT). Despite their capability in making statistically significant conclusions they may fail to capture exceedingly rare but devastating side effects, and the cases and time scale on which they are intended to be applied may not match what was tested in the trial.

3 Understanding Clinical Data

Health data will often come in tabular format, with many attributes such as test measurements and results for each patient at each clinic visit. However, it is difficult to observe or understand important patterns in the data from the "spreadsheet" view alone, and taking care and effort to visualize the data in another manner is especially important.

For instance, a histogram plotted for a numeric attribute in a dataset helps illustrate the distribution of that attribute and identify specific peculiarities to the data. One particular example shown in class is for the measured heart rate across patients in the MIMIC-III dataset, showing a bimodal distribution with a larger peak centered around a standard adult resting heart rate around

70bpm, and a smaller peak of high heart rate centered around 150bpm. Further inspection of the data and general health knowledge will allow us to understand that this second peak consists of infants, who have a much higher resting heart rate than adults and adolescents. Plotting the data separately for infants and adults by filtering on the age attribute allows us to confirm this, and deal with the data accordingly as needed. For instance, the "average" heart rate across the entire dataset may not be indicative of any standard individual as a large number of newborns in the data skews this average, and in some cases, depending on your goals, it may be more appropriate to split the data and perform different tasks on the separate groups.

Censored data When dealing with a health dataset, one must also make sure to properly handle censored data, which is likely to be present in datasets obtained from external sources for research purposes, given the private and sensitive nature of health data. For instance, to avoid personally identifying information for patients over the age of 90, the MIMIC-III clinical database records their ages as over 300. Thus, it is important that any analysis we do on this data does not rely on the specific numeric value given as the patient's true age, but perhaps we can group all these patients in a bin and denote their age with a specified categorical value higher than any other value.

Or, one may argue it could also be reasonable to record all such patients as age 90 as a numerical value for analysis, since each additional year past 90 is increasingly unlikely to be survived to and the total deviation in the dataset from the patient's true age will be as low as possible given what we know, though we also must acknowledge any simplifying assumptions made. In any case, an understanding of the data is essential in order to justify decisions made when using the data in any modeling procedures and in drawing broader conclusions from the data and its use.

4 Conclusions

Overall, when your aim is to help improve an aspect of the healthcare process with the use of health data, one must understand the task being done and ensure that it is coherent and well-defined, and understand the aspect that it aims to improve. One must also have a strong understanding of the specific dataset being used and do necessary pre-processing work in order to understand and interpret the use of any models being trained on this data, rather than just running the raw data on some off-the-shelf model. These notes give a summary of the course lecture given on this topic and provide a brief overview of the specific health tasks that can be done and are well-understood, as well as the importance of understanding and closely examining the data being used, through a few examples of possible peculiarities that may arise in real health data.