**Progress report – Dec 2019**

**Updated title:** Analysis of hospital based ayurvedic clinical practice to gain real world data knowledge

**Old Title:** Observational analysis of ayurvedic principles, ayurvedic hospital data, and patient outcomes

By Vinay Mahajan, Girish Tillu, Ashwini Mathur, Ashwini Godbole

Summary: The following progress has been made so far between July 2019 and Dec 2019

All the work done so far has been written as 5 chapters so far. Refinement and updates are expected as the research work continues.

Review of Real World Evidence methods and corresponding analysis.

EHR or EMR analysis using variety of methods:

When a patient visits a doctor or a hospital, what comes to anyone’s mind: “what is wrong?” and “what happens next?” The first question suggests the diagnosis of the illness; the other part is about prediction of future medical risk. At least till late 1990s, these questions were answered as well as solved by individual doctors. Year 2000 or so onwards, rapid growth in use of Electronic Medical Records (EMRs) or Electronic Health Records (EHRs) offers promises for healthcare analytics [Hillestad et al., 2005], such as chronic disease management and personalized medicine.

EMRs contain a wealth of healthcare information, including medication, procedure, diagnosis codes and lab test results. EHRs have created large amounts of data in different pockets of the world.

Bare minimum version of any EHR should include the following data points:

* Patient ID
* Patient gender
* Patient age
* Visit date
* Nature of visit (Inpatient / Outpatient)
* Disease diagnosis(s) coded in one of the many standard dictionaries
* Prescribed medicine(s) coded in one of the many standard dictionaries
* Administered procedure(s) coded in one of the many standard dictionaries
* Primary / secondary nature of the reported disease
* …

The IAIM hospital case report form contain all of the above mentioned data points plus a few more for the ayurvedic parameters.

In ayurvedic context where there are many practicing doctors writing this valuable only on case sheets, this large and rich data is not readily available for any kind of learnings.

EHRs, are usually heterogeneous, longitudinal, and sparse in nature. For each visit for each patient an EHR is filled, hence they become longitudinal. They contain the characteristics of temporal-sensitivity due to longitudinal nature of data. Temporal information is important to medical events, and is important to understand the dynamics of medical expressions. EHRs include a wide range of data (such as diagnosis, medication, lab test) and a large number of possible medical events (over ten thousands of diseases and medications), so that EHRs data is usually represented in a high dimensional space. EHR data is very sparse, since a record exists if and only if the patient pays a visit to a specific clinical institute, for a particular condition.

(i) Long-term dependencies in healthcare: the future illness and care may depend critically on historical illness and interventions. For example, the onset of diabetes at middle age remains a risk factor for the rest of the life; cancers may recur after years; and a previous surgery may prevent certain future interventions. (ii) Representation 1 arXiv:1602.00357v2 [stat.ML] 10 Apr 2017 of admission: an admission episode consists of a variable-size discrete set containing diagnoses and interventions. (iii) Episodic recording and irregular timing: medical records vary greatly in length, are inherently episodic in nature and irregular in time [6]. The data is episodic because it is only recorded when the patient visits hospital and is undergone an episode of care. The episode is often tightly packed in a short period, typically ranging from a day to two weeks. The timing of arrivals is largely random. (iv) Confounding interactions between disease progression and intervention: medical records are a mixture of the course of illness, the developmental and the intervening processes. In addition to addressing these four challenges, a predictive system should be end-to-end and generic so that it can be deployed on different hospital implementations of EMRs. An end-to-end system requires minimal or no feature engineering to read medical records, infer present illness states and predict future outcomes. [<https://arxiv.org/pdf/1602.00357.pdf>]

CoMorbidity analysis using “CoMorbidity” package in R.