Learning Effective Clinical Treatment Pathways from Observational Data
Rohit Vashisht, PhD¹, Kenneth Jung, PhD¹, Juan Banda, PhD¹ and Nigam H. Shah, MBBS, PhD¹.
¹Center for Biomedical Informatics Research, Stanford University School of Medicine, Stanford, CA, United States

Abstract
Treatment guidelines for control and management of type-2 diabetes mellitus (T2DM) remain controversial. Evidence from randomized clinical trials do not address many important clinical questions and are limited in their generalizability by exclusion criteria. Multiple treatment guidelines for T2DM suggest Metformin as first line medication, while the choices of second line drug remains ambiguous. A study by the Observational Health Data Science Initiative (OHDSI) found considerable variation between recommended guidelines and actual practice in T2DM. The factors underlying these variations, the effectiveness of a given treatment pathway and the best second line treatment for T2DM were not examined in this study. We used data from Stanford Hospital Electronic Health Record (EHR) to perform such an examination. Clinical features associated with the initial choice of treatment were re-discovered using a machine learning approach. Factors such as acute kidney disorder and liver disorder were predictive of first line therapy choices, thus correctly re-learning known guideline recommendations. In addition, the efficacy of first and second line treatments were evaluated using Cox proportional hazard models for control of Hemoglobin A1c on matched cohorts. DPP4-Inhibitor was found to be the most effective second-line therapy, and as effective as Biguanide as a first line therapy. Our approach, when implemented across the OHDSI network, could be an important step towards a learning healthcare system for informed medical decision making.

Introduction
Type-2 diabetes mellitus (T2DM) affects an estimated 29.1 million people in the United States [1]. Its global prevalence is projected to reach 440 million adults by the end of 2030 [1]. Current treatment guidelines, which are derived from a few randomized controlled trials [2-3], recommend the use of metformin (biguanide) as first line mono-therapy [4]. However, when metformin exhibits adverse effects or fails to control diabetes, a second line therapy must be chosen. There is little consensus on how to choose a second line therapy, with the American Diabetes Association recommending sulfonylureas, meglitinide (glinides), pioglitazone (thiazolidinediones) or dipeptidyl peptidase 4 inhibitor (DPP4) as second-line agent [5], and the American Association of Clinical Endocrinologists recommending alpha-glucose inhibitors, DPP4-inhibitors and GLP-1 agonist [6]. Given the availability of myriad treatment options for second-line therapy, and the availability of large amounts of EHR data, selecting an optimal second-line agent may be feasible using knowledge captured during routine clinical care. Thus
enabling learning health systems that can provide evidence for medical decision-making beyond that from formal clinical studies [7].

A recent study led by the Observational Health Data Science and Informatics initiative revealed significant diversity in the choice of first line therapy for T2DM [8]. Harnessing data from 11 databases that collate 250 million records into a unified common data model, the study found that metformin was the predominant initial choice of therapy but that other choices were also common. Substantial heterogeneity in the prescription of second-line agents was also noted, highlighting a gap in available clinical guidelines for management of T2DM. Important clinical questions such as the factors that determine the initial choices of treatment, and the comparative effectiveness of second-line therapies were not addressed. In pursuit of these goals, we set out to perform a systematic analysis of treatment decisions in T2DM using data collated in Stanford’s clinical data warehouse.

**Conclusion**

We recapitulate previous work regarding variation in the choice of first line therapy, and find clinical factors that are predictive of the first line therapy choice that are consistent with biomedical knowledge of adverse effects associated with metformin. Finally, we demonstrate the feasibility of comparative effectiveness studies of second line therapies in controlling HbA1c using matched cohorts that adjust for comorbidities that might impact the treatment outcome. DPP4-Inhibitors appears to be as effective as metformin as a first line therapy, and is considerably better than other options as a second line therapy. At present, our analysis is limited to data from Stanford Hospital, but could be extended to any site that has adopted the OHDSI common data model.

**References**

5. Association AD. Standards of Medical Care in Diabetes—2016. Diabetes Care 2016;39(Supplement 1)