**Medical Record Data to augment understanding of pediatric to adult care transition readiness**

**of adolescents and young adults with chronic illness**

For the summer BD2K training program we propose an interdisciplinary project about pediatric to adult care transition readiness and adherence to medical treatment using medical record "big data". Below is a description of our project in a little more detail.

Dr. Ferris, the leader of our research team, is a world-renowned leader in understanding transition of adolescents and young adults (AYAs) with chronic conditions from pediatric care settings to adult care settings. Her team has been following over 600 patients at NC Children’s Hospital for 10 years, and measuring their knowledge and self-management skills as they age. Recently, we have obtained the complete medical care history for those patients from the medical record system.

With this big data set, we attempt to answer two important and interesting research questions that have not been addressed in previous studies. First, we intend to quantify the determinants of disease knowledge and self-management proficiency using econometric and applied statistics methods. Determinants include factors beyond the patient’s control (such as gender, race, parent’s education, years since disease onset) as well as, and importantly, factors for which the patient is responsible (such as appointment adherence and medication adherence). In order to measure the unbiased impact of these latter determinants on disease knowledge and self-management (which define readiness for successful transition and future health outcomes), we plan to jointly estimate a dynamic model of the knowledge production function (which measures time-varying impacts of factors on knowledge gains over time) and the medical care visit and medication adherence behaviors of the patient between observations over time. The dynamic model allows us to evaluate the role of adherence on measures of transition readiness. It also allows us to measure the effect of knowledge and self-management skills on subsequent adherence. Second, we try to quantify the determinants of patients' adherence to medical treatment, including their knowledge and self-management skills. With the rich and complete information about patients' medical care behaviors over time, we are able to apply some "big data" techniques such as machine learning to stratify the sample and let the data tell us the most suitable economic behavioral model for each group of people.

We use the provider-verified and disease-neutral *TRxANSITION Scale* scores as the measure of patient knowledge and self-management skills. It has ten subsections that evaluate a patient's knowledge about his/her disease, nutrition, medication, as well as self-managing skills. A total of 32 questions are asked by a medical provider and/or research assistant knowledgeable of the AYA's condition (either in person or over the phone). During 2006 to 2015, 638 patients between the ages of 12 and 31 (inclusive) who have been diagnosed with a chronic health condition were recruited from UNC hospitals, followed annually over time, and included in our sample.

Patients' whole histories of medical care information, including their diagnoses, clinic appointments, emergency room visits, treatments, lab measures, and medications in the UNC hospitals have been scraped from the medical records by NCtraCS. Basic demographic, socioeconomic, and educational information of the patients is also collected from the interview or medical care record system. A variety of diseases are included in our sample, including chronic kidney diseases, diabetes, hypertension, inflammatory bowel disease, systemic lupus erythematosus, sickle cell, genetic condition, Leukemia, and HIV.

There are several places that computer science, Statistics, Applied Mathematics, and medical science students can play an important role in the project with the skills they acquired in their training. First, the medical care data set is well organized but in different styles for each type of information. For example, procedure information is at the "encounter" level, vitals and lab measures are at the patient-time level, and charge data are at the patient-procedure level. CS students can apply their data processing and managing skills to most efficiently combine and clean these millions of observations in each data set and medical science students can provide useful suggestions about how to deal with information we have given their background in medical care training. Furthermore, this project provides an opportunity for students in different fields to work with medical record data together, and may bring innovative suggestions for medical record data use in the future. Second, the model estimation part requires thorough statistical analysis, where statistics and applied mathematics students can provide guidance. With the quantitative and economic modeling skills acquired through Master’s and Ph.D. training to date of the economist on our team, we can work together to improve the methodology in both the fields of economics and medical care. In addition, we can apply methods and techniques in "big data" analysis to improve the interdisciplinary perspective.

The diversified mentor group for the project includes:

Donna Gilleskie, PhD, Department of Economics, UNC-CH

(Donna is an expert in health economics and applied econometrics.)

Maria Ferris, MD, PhD, School of Medicine, UNC-CH

(Maria is an expert in pediatric to adult care transition and pediatric nephrology.)

Miranda A.L. van Tilburg, PhD, Department of Gastroenterology and Hepatology School of Medicine, UNC-CH

(Miranda is an expert in pediatric research, especially in the field of pediatric functional gastrointestinal disorders.)

Eniko Rak, PhD, School of Allied Health, UNC-CH

(Eniko is an expert in clinical rehabilitation and mental health.)

Specifically, our needs include the following:

* Initially, we desire help with coding all the available information on patient medical care utilization into a usable format.
	+ Dated patient level encounters (scheduled physician visits, outpatient visits, inpatient stays, ER visits) [used to construct both adherence to scheduled visit, as well as unexpected shocks to health]
	+ Dated patient level prescriptions and refills [used to construct measure of adherence to taking prescribed meds to treat illness]
	+ Dated charge information at procedure level [might as well clean these even though not a part of our current project]
	+ Dated patient level lab results on important markers of disease progression
* Then, we will use these data to supplement our “knowledge production function analysis” where knowledge gains (or losses) are modeled over time as a function of demographic, educational, parental, and disease characteristics as well as endogenous encounters and adherence.
* Finally, there is an underlying disease process that we can obtain from the lab values.  Patient makes decisions (about encounters and adherence) not knowing the true state of disease but rather having some expectation (or perhaps knowledge of the distribution) of the disease state.  We’d like to jointly model evolution of the disease state, knowledge of patients, adherence. and health shocks.  (As econometricians who estimate such dynamic multiple equation models all the time, we have the skill set to do this part, but I include it in case this is an area where you can learn our tools.)