## 1.1 Background

From 1990 to 2010, the number of people with a diabetes diagnosis more than tripled, from 6.5 million to 20.7 million.1 As of 2015, approximately 9.4% of the United States population have diabetes.2 The increased burden of diabetes bears large costs to society; the American Diabetes Association estimates the direct cost of diabetes at $237 billion in 2017, or approximately 1 in 4 health care dollars spent in the United States.3

There exist wide-ranging disparities in diabetes prevalence,4,5 quality of care,6,7 and outcomes8–10 in the United States. In terms of race, American Indians, Black, and Hispanic patients account for a disproportionate share of diabetes complications and worse disease-related outcomes,11,12 whereas Whites have a higher risk of all-cause mortality and cardiovascular disease compared to ethnic minorities. Prevalence of diabetes has significantly increased in both White and Black residents of the southeastern Stroke Belt states, indicating regional variation.5 Further, diabetes is more prevalent and inadequately managed in rural areas.13,14 Recent cost-saving trends towards high deductible health insurance plans disproportionately impact lower-income individuals, who may forego necessary care until the disease progresses.15

Although studies reported an improvement in population achievement of diabetes treatment goals from 1990-2010 - recent data indicate there has not been an improvement from 2005 to 2016.16,17 Currently, only an estimated 23% of those with diabetes engaged in health care and met four major care goals: blood pressure, cholesterol, lipoprotein cholesterol target, and smoking abstinence.17 Clinicians may fail to escalate treatment to achieve treatment goals - even though patients are not reaching glycemic targets.18 Termed clinical inertia, this inefficient care delivery may also be influenced by a patient’s sociodemographic factors. Research suggests that older and White patients are more likely to have treatment intensified at lower HbA1c compared to younger and Black patients.19

Given the progressive nature of diabetes, quality care delivery along the care cascade— the process of diagnosis, linkage to care, and the achievement of treatment targets—is necessary to prevent the development of severe complications and comorbidities. Disparities exist along the diabetes care cascade, as young adults, women, non-Hispanic Blacks, and patients that were covered by Medicaid or uninsured are less likely to meet care goals.17,20

Although interventions have been conducted to address clinical inertia and improve provider behavior,21 there remains gaps in the literature in describing how health care use varies among those with diabetes. This analysis seeks to understand those utilization patterns and identify whether trends in health care use vary by sociodemographic groups.

## 1.2 Rationale

Although trends in outpatient use, ED visits, and hospital discharges have been examined by race, age group, sex, complication type, and health insurance coverage, there has been no data published on trends by geographic region, state, and rural/urban location.22,23 Further, the aforementioned data describe trends until 2011. This provides an opportunity to both update and further describe healthcare utilization trends among people with diabetes.

Further, existing data show distributions of healthcare use, but there are no data examining the same people linked across datasets and how they use health care each year. Using data that link individuals throughout the continuum of care, we can examine what differentiates individuals that are readmitted and those that are not.

Methods

The analysis will be conducted using a retrospective serial cross-sectional design using data from the AHRQ’s Healthcare Cost and Utilization Project. Specifically, data from the National Inpatient Sample, the Nationwide Emergency Department Sample, and the State Inpatient Database, and the State Emergency Department Database.

The AHRQ’s Healthcare Cost and Utilization Project (HCUP) is the “largest collection of all payer, encounter-level hospital care data in the United States.” There are multiple HCUP datasets: The National Inpatient Sample (NIS), the Nationwide Emergency Department Sample (NEDS), the State Inpatient Database (SID), and the State Emergency Department Database (SEDD). Each dataset contains hospital-level claims data.

The National Inpatient Sample is a database of hospital inpatient stays derived from billing data by U.S community hospitals. Data are systematically sampled from the State Inpatient Databases. Each year of the NIS includes over 7 million inpatient stays. The Nationwide Emergency Department Sample contains data from approximately 31 million ED visits per year and estimates roughly 143 million ED visits. These datasets are available for purchase through the HCUP website. The HCUP data use agreement requires that researchers do not attempt to discover the individual identity of anyone in the database

Patients with diabetes will be identified by the presence of a diabetes-specific ICD-9 or ICD-10 code. Variables for age, race/ethnicity, geographic region, urban/rural location, and insurance type are each included in the SID, SED, NIS, and NEDD. We aim to capture the health care utilization of the non-institutionalized population aged 18+ which are diagnosed with diabetes as indicated by presence of a diabetes-specific ICD-9 or ICD-10 codes which visited hospital inpatient and emergency department settings from years 2005-2016. We will stratify our results by age, race/ethnic group, sex, health insurance coverage, geographic region, and urban/rural designation. Results will be produced at both the national and state level.

Validation of the data will be run alongside the descriptive analysis, and data that are missing or incorrect will be corrected if possible and otherwise excluded. If analysis indicates that the missing data are missing completely at random, case deletion will be used to progress the analysis. Otherwise, multiple imputation methods will be used to impute values. The specific imputation model used will depend on the type of data element in the analysis that require imputation.

During 1988 to 2011, the NIS was constructed annually by including 100% of the discharges from 20% of US hospitals. Starting in 2012, the AHRQ redesigned the NIS as a 20% national patient-level sample in 2012, with nonrepresentative sampling across hospitals.46 This change will be addressed through the use of trend weights which allow for trend analysis prior to and post sampling change.47

The analysis will use the Taylor-series linearization and/or the jack-knife methods to estimate the standard errors associated with weighted estimates.

Rate estimates will be calculated as follows:

𝑇𝑜𝑡𝑎𝑙 𝐻𝑒𝑎𝑙𝑡ℎ𝑐𝑎𝑟𝑒 𝑆𝑒𝑟𝑣𝑖𝑐𝑒 𝑈𝑠𝑒∗ 𝑖𝑛 𝑋𝑋 𝑦𝑒𝑎𝑟 × 1000

𝑁𝑢𝑚𝑏𝑒𝑟 𝑜𝑓 𝑃𝑒𝑜𝑝𝑙𝑒 𝑤𝑖𝑡ℎ 𝐷𝑖𝑎𝑏𝑒𝑡𝑒𝑠 𝑖𝑛 𝑋𝑋 𝑦𝑒𝑎𝑟

\*Where Healthcare Service Use is defined as number of Hospital Inpatient Stays,

ED visits, Potentially Preventable Hospitalizations, and Drugs Prescribed

In order to create a denominator for use in estimating rates per 1000 persons with diabetes, National Health Interview Survey data will be used to determine the population of people with diabetes. The population estimates generated will use self-reported diabetes diagnosis to indicate presence of diabetes disease state. NHIS data will be also be used to estimate the population of people with diabetes by sex, race/ethnicity, rural/urban designation, geographic region, and health insurance coverage status. These estimates will be used to generate rates per stratification of interest.

Percent change estimates will be calculated as follows:

𝐴𝑏𝑠𝑜𝑙𝑢𝑡𝑒 𝐶ℎ𝑎𝑛𝑔𝑒 𝑓𝑟𝑜𝑚 2008 − 2016

× 100

𝐸𝑠𝑡𝑖𝑚𝑎𝑡𝑒 𝑓𝑜𝑟 𝐻𝑒𝑎𝑙𝑡ℎ𝑐𝑎𝑟𝑒 𝑆𝑒𝑟𝑣𝑖𝑐𝑒 𝑢𝑠𝑒 𝑖𝑛 2008

Self-reported diabetes diagnosis will be used to indicate presence of diabetes disease state. In order to account for the complex survey design of the National Health Interview Survey, a sample weight will be applied to generate these estimates.

Each analytic procedure/code will first be conducted and validated with one year of data, and then applied and adjusted for use with prior years after confirmation of validity.

Continuous variables will be expressed as the mean ± standard deviation. Categorical variables will be presented in terms of expected count and frequency in the dataset. Rates will be expressed per 1000 persons with diabetes.

No statistical testing of differences will be performed as the analysis was designed as a descriptive study.

All statistical analysis will be performed using R (R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org/).51