SEARCH for Diabetes in Children and Young Adults 0-45 years (SEARCH-DiCAYA) Diabetes Surveillance Study

Protocol

September 30, 2020 - September 29, 2025

Version 1: December 5, 2020

Version 2: May 10, 2021

Version 3: October 27, 2021

Version 4: October 27, 2022

Version 5: February 15, 2023

Table of Contents

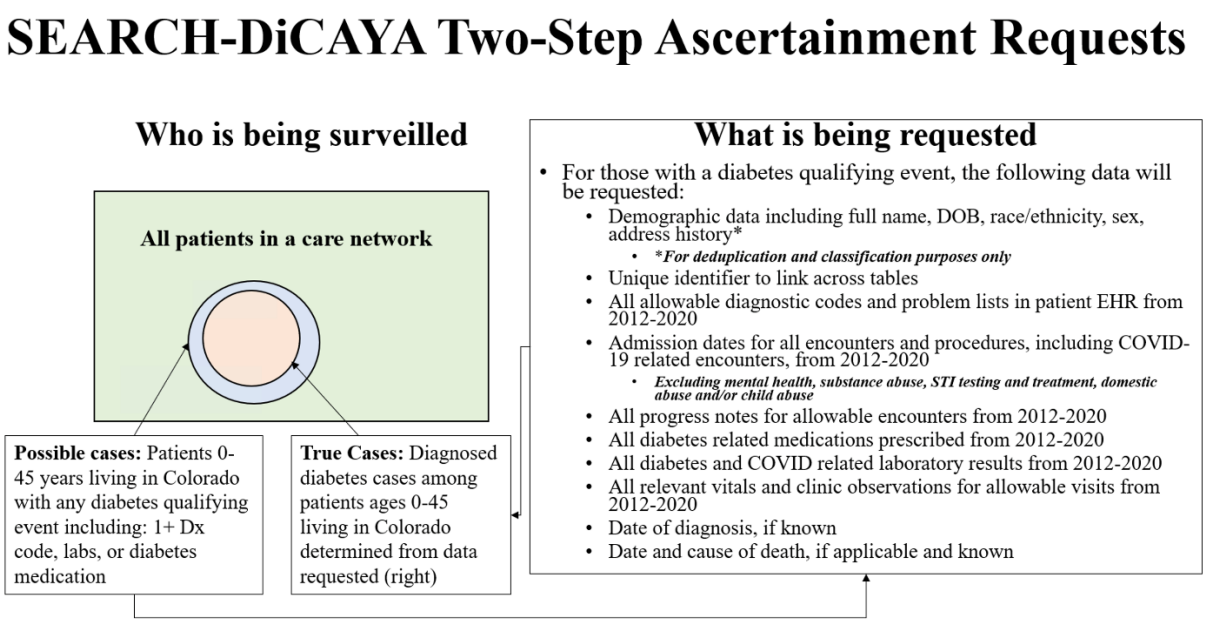
SEARCH-DiCAYA Diabetes Surveillance Protocol

I. HYPOTHESES AND SPECIFIC AIMS

Diabetes is one of the most common chronic diseases of childhood, estimated by the SEARCH study to impact 2.22 in every 1,000 youth aged <20 or ~3.3 million youth in the U.S. (1) Findings from SEARCH suggest that the incidence and prevalence of both type 1 (T1D) and type 2 diabetes (T2D) are increasing among U.S. youth.(2, 3) Between 2001-2009, the prevalence of T1D increased by 30.0% from 1.48 to 1.93 per 1000.(2) While, historically T1D has been considered to effect primarily white youth, increases in prevalence are seen in all ages and race/ethnic groups except American Indians.(2) The adjusted incidence of T1D increased in the SEARCH study by 1.4% annually between 2002-2012 with the steepest increase observed among Hispanic youths (annual increase, 4.2%).(3) Between 2001 and 2009, the prevalence of T2D in the SEARCH population increased by 30.5% overall with significant increased trends observed in all race/ethnic groups except American Indians and Asian or Pacific Islanders, possibly due to smaller sample sizes.(2) Similarly, the incidence of T2D was found to increase at an annual adjusted rate of 4.8% between 2002-2012 with significant annual increases among all race/ethnic groups except non-Hispanic whites.(3)

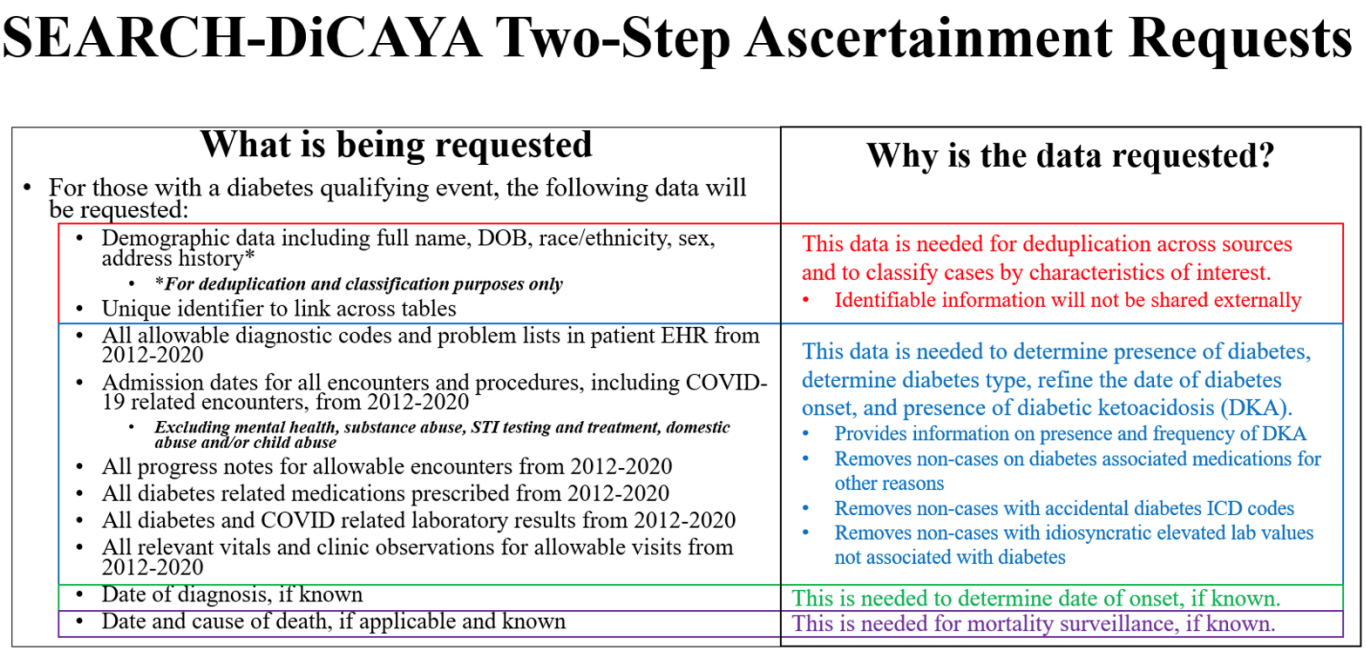
In the SEARCH for Diabetes in Children and Young Adults study (SEARCH-DiCAYA), we aim to **ascertain the annual prevalence and incidence of diabetes among youth and young adults <46 years of age in the state of Colorado starting with year 2020** from the well-established network of pediatric and adult endocrinology clinics, community clinics and hospital networks in Colorado during SEARCH and expand into a broader network including additional data sources into the new enhanced SEARCH-DiCAYA network designed to surveil diabetes trends in youth and young adults.

Figure 1. SEARCH-DiCAYA Two-Step Case Ascertainment Data Requests



Our integrated surveillance approach will utilize algorithms to identify diabetes cases, distinguish diabetes type, and estimate onset date in a two-step process where Colorado case ascertainment network sources will first identify possible cases of diabetes with one or more diabetes qualifying events (e.g.,- diabetes diagnostic code(s), prescription of diabetes-associated medication(s), and/or diabetes qualifying laboratory results) and then provide comprehensive datasets of EHR extracted data from 2012 onward on all algorithm-identified possible cases including: demographic data (PHI including full name, date of birth, sex, race/ethnicity, and address history for deduplication purposes), all diagnosis codes, encounters and procedures, all diabetes associated laboratory values, all diabetes associated medication prescriptions, all progress notes for applicable encounters, all relevant vitals and clinical observations, date of diabetes diagnosis, and date and cause of death; if applicable.

Figure 2. SEARCH-DiCAYA Two-Step Case Ascertainment Data Request Purpose of EHR Data Requested



Incremental record linkage will be used to minimize missing data on key surveillance variables (i.e., race/ethnicity, date of diabetes diagnoses, diabetes type) and improve data quality by integrating encounter information across data sources. Targeted chart review will be conducted on categories of poor predicted performance (type 2 cases and other diabetes cases) and limited validation to generate timely estimates with high overall accuracy and validity. The breadth of EHR data requested on possible cases is needed to reduce the amount of chart review needed on cases where the presence, type, or date of diabetes diagnosis are unclear. Data going back to 2012 is requested to ensure that correct assignment of onset (incident) year is made to assist with discerning new onset (incident) cases from longer term existing (prevalent) cases particularly for those who receive healthcare sporadically and/or across multiple sources.

Aim 1: PREVALENCE – Starting in year 2020 and continuing through year 2024, to ascertain prevalent diabetes cases among youth 0-17 years of age and adults 18-45 years of age on a yearly basis from an integrated surveillance system using algorithms designed to distinguish diabetes type.

Research question 1.1: What are the yearly state-level age-, sex- and race/ethnicity-specific prevalence estimates of T1D and T2D among youth and young adults in Colorado?

Research question 1.2: Building off prevalence data from SEARCH-Colorado, what are the long-term temporal trends in the prevalence of T1D and T2D and do they vary by race/ethnicity, age, and sex?

Research question 1.3: Has the COVID-19 pandemic (2020-present) altered trajectories of temporal trends in the prevalence of T1D and T2D and, if so, do they vary by race/ethnicity, age, and sex?

Research question 1.3b: Does COVID-19 associated diabetes prevalence affect temporal trends in diabetes associated vs. non-diabetes associated effects on premature morbidity and mortality for youth and young adults with diabetes?

Aim 2: INCIDENCE – Starting in year 2020 and continuing through year 2024, to ascertain newly diagnosed diabetes among youth 0-17 years of age and adults 18-45 years of age on a yearly basis from an integrated surveillance system using algorithms that estimate date of onset and distinguish diabetes type.

Research question 2.1: What are the yearly state-level age-, sex- and race/ethnicity-specific incidence estimates of T1D and T2D among youth and young adults?

Research question 2.2: Building off incidence data from SEARCH-Colorado, what are the long-term temporal trends in the incidence of T1D and T2D and do they vary by race/ethnicity, age and sex?

Research question 2.3: Has the COVID-19 pandemic (2020-present) altered trajectories of temporal trends in the incidence of T1D and T2D and, if so, do they vary by race/ethnicity, age, and sex?

Research question 2.4: What is the prevalence of diabetic ketoacidosis (DKA) at the time of diagnosis youth and young adults with T1D and T2D?

Research question 2.5: Building off data from SEARCH-Colorado, what are the long-term temporal trends in prevalence of DKA at diagnosis with T1D and T2D?

Aim 3: EVALUATE PUBLIC HEALTH SURVEILLANCE METHOD – To evaluate the strengths and challenges of our integrated surveillance approach to determine the burden of diabetes among youth 0-17 years and adults 18-45 years by ascertaining validity, completeness and representativeness of case ascertainment methods.

Research question 3.1: What is the sensitivity, specificity, and positive predictive value of the SEARCH-DiCAYA surveillance approach, overall and by diabetes type in a random sample of T1 and T2 cases?

Research question 3.2: What are the capture-recapture adjusted prevalence and incidence estimates of T1 and T2 diabetes while accounting for the potential bias introduced by less than complete case ascertainment?

II. BACKGROUND AND SIGNIFICANCE

Diabetes is one of the most common chronic diseases of childhood in the United States (U.S.) and global trends suggest the condition is increasing in incidence worldwide.(4-7) With slower progress in diabetes prevention than was hoped, the number of children being diagnosed with diabetes has important implications for the planning and delivery of health services. Childhood onset of diabetes results in a longer duration of disease in which complications are duration dependent, thus persons diagnosed in childhood will likely face chronic kidney disease and dialysis (8, 9), coronary heart disease and cerebrovascular disease at younger ages than those whose diabetes develops during adulthood.(10) Further, increasing incidence in both T1D and T2D in youth means that more women will have pregnancies complicated by diabetes, increasing the risk of obesity and diabetes in their offspring.(11, 12)

The Overall Burden of Diabetes among Youth in the U.S. In the youth population in the U.S., the prevalence of all diabetes types increases with age and there are marked disparities by race/ethnicity. In the SEARCH study the crude prevalence of **all diabetes types** was 2.22 per 1,000 youth aged <20.(1) T1D accounted for 89% of the prevalent diabetes cases in SEARCH in 2009 and 98% of the cases were among children <10 years of age.(1) Total diabetes prevalence in youth increased by age group from 0.30 per 1,000 youth aged <5 years, 1.4 per 1,000 youth aged 5-9, 3.0 per youth aged 10-14 and 4.0 per 1,000 youth aged 15–19 years.(1) Overall, non-Hispanic white (NHW) youth had a higher prevalence of any diabetes type in 2009 (2.6 per 1,000 youth aged <20) followed by non-Hispanic black (NHB) at 1.6 per 1,000, Hispanic of any race (1.29 per 1,000), Asian and Pacific Islanders (ASPI) at 0.6 per 1,000 and American Indian and Alaska Natives (AIAN at 0.35 per 1,000.(1) Estimates of the prevalence of diabetes from the 2017-18 National Survey of Children's Health (NSCH) (13), based on parental self-report, estimated the prevalence of all types of diabetes at 4.0 per 1,000 children <17 years of age. Variation was observed by race/ethnicity with the highest prevalence among NHB youth (12 per 1,000) followed by NHW youth (3.0 per 1,000) and the lowest among Hispanic youth (2.0 per 1,000). When comparing HRSA Regions in the U.S., the highest prevalence of diabetes from the NSCH is seen in HRSA Region 6 (New Mexico, Texas, Oklahoma, Arkansas, and Louisiana) at 9.0 per 1,000 and the lowest is seen in Region 9 (Nevada, California, Arizona, and Hawaii) at 1.0 per 1,000. Colorado is located in HRSA Region 8 where the prevalence of diabetes was 3.0 per 1,000.

The Burden of Type 1 Diabetes in Youth. T1D was estimated by SEARCH to affect 1.93 per 1,000 youth <20 years of age in the U.S. in 2009.(1) The highest prevalence was seen in NHW youth at 2.6 per 1,000, followed by NHB and Hispanic youth at 1.6 and 1.3 per 1,000, respectively.(1) Between 2001-2009, the prevalence of T1D increased from 1.48 to 1.93 per 1000, representing an increase of 30.0%. (2) While, historically T1D has been considered to affect primarily white youth, increases in prevalence are seen in all ages and race/ethnic groups except AIAN.(2) The increase in prevalence seen among minority race/ethnicity groups in the

U.S. is of concern because the SEARCH study reported that minority youth have poorer glycemic control, a major risk factor for diabetes complications.(14)

Incidence of Type 1 Diabetes in Youth. A significant upward trend has also been observed in the incidence of T1D in the SEARCH population, from 19.5 cases per 100,000 youths per year in 2002–2003 to 21.7 cases per 100,000 in 2011–2012; an annual increase of 1.4%.(3) The adjusted rise in the incidence of T1D was especially pronounced among Hispanic youths, among whom the annual increase was 4.2% compared to 1.2% among NHW youths ($p<0.001$). The incidence increased in the 5-9 and 15-19 year age groups and among boys but not girls.(3) The trends in incidence observed in the SEARCH study in the U.S. are similar to those seem in Europe. While the overall rate of increase in the EURODIAB study suggested a plateauing between 2004-2008, the incidence rates returned to previous levels in 2009-2013, consistent with a previously proposed cyclic effect.(4) Rates in Europe increased at similar proportions among boys and girls in the 0-4 year age group (3.7% and 3.7% annually) and in the 5-9 year age group (3.4 and 3.7% annually). However, the rate of increase was higher among boys in the 10-14 year age group (3.3% and 2.6% annually).(4)

The Burden of Type 2 Diabetes in Youth. The prevalence of T2D among youth in the SEARCH study in 2009 was estimated to be 0.24 per 1,000 youth aged <20 years.(1) The highest burden was observed among AIAN (0.63 per 1,000), NHB (0.56 per 1,000) and Hispanic youth (0.40 per 1,000), while the lowest burden was noted NHW youth (0.09 per 1,000), a pattern that is almost the inverse of that seen in T1D.(1) Between 2001 and 2009, the prevalence of T2D in the SEARCH population increased by30.5% overall with significant increased trends observed in all race/ethnic groups except AIAN and ASPI, possibly due to smaller sample sizes.(2)

Incidence of Type 2 Diabetes in Youth. The incidence of T2D in the SEARCH study was found to increase at an annual adjusted rate of 4.8% between 2002-2012 with significant annual increases among all race/ethnic groups except NHW youth.(3) It is hypothesized that variations in the underlying prevalence of obesity in subpopulations of race/ethnicity and sex over time may contribute to the heterogeneity in incidence trends of type 2 diabetes.(15) The observed annual rate of increase of 4.8% is higher than modeled by Imperatore et al.(16) (2.3%) in which the prevalence was estimated to rise by 178% by 2050 to 0.75 per 1,000 youth. If current trends in incidence continue, it is projected that the number of youth with T2D will double by 2050, primarily due to shifting U.S. demographics.(16)

Changing Clinical Presentation. Diabetic ketoacidosis (DKA) is a serious, life-threatening acute complication of diabetes and most commonly occurs at the time of diagnosis.(17) DKA prevalence among youth with T1D and T2D in SEARCH was 31.1% and 5.7%, respectively in 2008-2010.(18) However across the period 2010-2016, the SEARCH study observed a 2% annual increase in the prevalence of DKA at diagnosis of youth-onset T1D at 38.5% ((19), manuscript submitted). This represents a ~31% relative increase since 2002-2010 which disproportionately impacted younger children, racial/ethnic minorities, and children from uninsured families, however, was not explained by changes in the distribution of sociodemographic factors among incident T1D cases over time. This is of considerable concern because DKA is associated with acute complications, increased mortality (20) as well as poor future glycemic control.(21)

Challenges in Classifying Diabetes Type in Youth. Although T1D remains the predominant form of diabetes diagnosed during childhood, T2D is emerging as a serious pediatric condition. T1D is characterized by profound hyperglycemia due to absolute insulin deficiency caused by

immune associated destruction of the insulin-producing beta cells of the pancreas. Nearly 90% of individuals with T1D have presence of one or more islet cell autoantibodies such as insulin (IAA), glutamic acid decarboxylase (GADA), insulinoma- associated autoantigen 2 (IA-2), and zinc transporter 8 (ZnT8A) at some point in time, especially during the pre-clinical stages of disease.(22) However, diabetes autoantibody testing is not routinely done or easily accessible for surveillance purposes. The pathogenesis of T2D in youth involves dual defects of insulin resistance and beta-cell dysfunction and the disease is more aggressive than in adults with a faster deterioration of beta-cell function(23) and poorer response to glucose-lowering medications.(24) The physiologic framework developed by the American Diabetes Association (ADA) in 1997 and updated in 2010 recommends classifying diabetes type into 3 broad categories: type 1, evidence of autoimmune-mediated beta cell destruction usually leading to absolute insulin deficiency; type 2, a combination of insulin resistance and an inadequate compensatory insulin secretory response; and other specific types including genetic defects of beta-cell function or insulin action, diseases of the exocrine pancreas, endocrinopathies and drug, chemical or infection induced diabetes.(25) However, a variety of challenges makes phenotyping the presence and type of diabetes among youth for surveillance purposes difficult. Determination of evidence for autoimmune- mediated beta-cell destruction is difficult as diabetes autoantibody testing is not routinely performed in clinical care or easily accessible for surveillance purposes. Further, beta-cell autoantibodies disappear with time and might even be absent at the time of T1D diagnosis.(26) Insulin secretion tests are difficult to perform and interpret and the clinical characteristics to distinguish diabetes type in youth are becoming less useful. The rising prevalence of obesity in the childhood population in the U.S.(15) minimizes the usefulness of body mass index as a distinguishing feature between type 1 and type 2 diabetes. Further, obesity in childhood may contribute to convergence of an insulin resistant phenotype of T1D that could be misdiagnosed based on clinical judgement alone.(27) Diabetic ketoacidosis (DKA) at clinical presentation of T1D in youth is common, however DKA can be present (though at a lower frequency) in youth with T2D.(18) A surveillance approach based on electronic medical record, insurance claims or other clinical databases must include thoughtful consideration to the various challenges of diagnosis of diabetes type in youth.

Challenges in Surveillance of Diabetes in Youth. In order to monitor trends of T1D and T2D in youth, efficient, flexible, and sustainable surveillance systems are needed to provide timely estimates of prevalence and incidence by sex, race/ethnicity and age group categories while maximizing data quality. The SEARCH surveillance approach (28) identified youth < 20 years of age with prevalent and incident diabetes since 2001 for prevalence and since 2002 for incidence from six clinical centers, four geographic-based sites and two health-plan based sites with ~5.5 million children <20 years of age (~6% of the U.S. population <20) under surveillance annually. (29) The approach resulted in consistently high estimates of completeness of case ascertainment over time for both incident cases [T1D: 98.5% to 98.9%; T2D: 91.6% to 94.0%] (3) and prevalent cases [T1D: 92.5% to 99.3%; T2D: 92.9% to 96.1%]. However, the SEARCH approach is costly, time- and labor-intensive with manual validation of diabetes status, diabetes type and date of diagnosis in all ascertained cases via chart review, and requires a long window of ascertainment (11.2 and 25.3 months to identify 90% of the incident T1D and T2D cases, respectively). (30) Thus, a new, similarly accurate, but more sustainable surveillance system to ascertain diabetes by type in youth and young adults is needed to generate timely estimates of prevalence and incidence in demographic subpopulations in the U.S. Over the past decade SEARCH researchers have harnessed the electronic health record (EHR), insurance claims or

other clinical databases to conduct surveillance of diabetes in the youth population using automated algorithms.(31-33) However, a variety of challenges are posed by reliance on existing clinical or administrative data which must be considered to ensure that prevalence and incidence estimates are not biased by incomplete case ascertainment, misclassification of diabetes, or external validity concerns.

Algorithms to Identify Diabetes in Youth and Young Adults from the EHR, Insurance Claims and Clinical Administrative Databases. EHR and insurance claims-based algorithms have generally been shown to perform well to identify diabetes cases in youth, but are challenged to distinguish between diabetes types, particularly an accurate ascertainment of youth with T2D. As part of the SEARCH study, Lawrence et al. (31) evaluated the performance of a variety of EHR-based algorithms to identify youth with diabetes, both overall and by diabetes type among enrollees <20 years of age in the Kaiser Permanente Southern California Health Management Organization, using the SEARCH protocol for validation. The algorithm with the best performance had a sensitivity of 95.9%, PPV of 95.5% and an accuracy of 97.9% to identify true SEARCH – cases. The algorithm to classify diabetes type preformed respectably for T1D (sensitivity=94.8%, positive predictive value=98.0%), however it was less rigorous in the correct identification of T2D (sensitivity = 93.2%, PPV=81.8%). Among youth from the University of North Carolina Health Care System, and also within the SEARCH context, Zhong VW, et al. (32) evaluated automated algorithms utilizing administrative payment and EHR data. Similarly, the authors found high sensitivity, PPV and specificity of their type-insensitive algorithm to identify true diabetes cases (based on the SEARCH protocol), but reported considerable differences in the performance of algorithms to distinguish type, particularly for T2D. Outside of SEARCH, in the National Diabetes Surveillance System in Canada, which uses administrative health data to identify diabetes in youth <20 years of age, algorithms that incorporate demographic data and drug utilization patterns had high sensitivity for T1D (98.6%) but misclassified T2D 17% of the time. (34) T2D in youth disproportionately impacts lower-income subpopulations which may be more likely to cross healthcare systems, thus resulting in incomplete clinical information in any one clinical system. **Thus, a surveillance approach is needed that identifies potential diabetes-related healthcare encounters across healthcare systems that serve the target population and an accurate record linkage solution to link care received by the same individual.**

Challenges Regarding the Linkage of Healthcare Received by Youth and Young Adults with Diabetes across Clinical and Administrative Datasets. Surveillance systems for chronic diseases in the U.S. are increasingly turning towards linkage of EHRs and medical claims to generate timely estimates of the rate and burden of disease. Population-level pediatric and adult diabetes surveillance systems that rely on EHR and/or insurance claims data in the U.S. are challenged by issues related to the fragmented health care system and variable data quality across clinical settings (e.g., missing data elements) and data sharing/privacy concerns.(35) The EHR is designed for clinical operations (36-38), rather than research use, thus data error issues (e.g., omissions and misclassifications) and imprecise mapping of diabetes diagnostic or laboratory results to diabetes phenotypes may result in difficult to interpret findings (39) and validity concerns when compared to findings from prospective cohorts.(40) Further, the population coverage in EHR data from a single healthcare system is inherently nonrandom and thus biased

by the institution's target care population, health insurance coverage, clinical services offered and socioeconomic factors.(41) Pediatric and adult diabetes surveillance systems that conduct case finding across diverse networks of healthcare systems offer the opportunity to identify a more representative population of cases from the underlying source population than could be detected in a single system. However, the fragmented nature of the U.S. healthcare system and the nonexistence of a universal patient identifier across systems necessitates semi-automated, accurate record linkage solutions. Individual- level health encounter data is often scattered across disparate healthcare systems, particularly for children and adults with T2D. (42) The proposed expanded SEARCH-DiCAYA surveillance system will identify potential cases with diabetes-related healthcare encounters in the 5 years preceding a given surveillance year from the SEARCH network of pediatric and adult endocrinology clinics, community clinics and hospital networks in Colorado and expand into a broader network including additional data sources into the new enhanced SEARCH-DiCAYA network. A semi- automated, hybrid deterministic and probabilistic record linkage method (which is designed to allow for variable data quality and missing data) will be used to identify encounters across disparate systems belonging to the same individual and create a timeline of diabetes-related healthcare encounters over the 5 year period.(43)

III. RESEARCH METHODS

To accomplish the specific aims of SEARCH-DiCAYA, there will be two primary activities: 1) ascertainment of prevalent and incident cases of diabetes among youth 0-17 years of age and young adults 18-45 years of age using semi-automated algorithms to determine diabetes status, diabetes type and estimate onset date with targeted chart review on cases that fall into categories of predicted poor performance, and 2) execution of protocols to understand the validity of our surveillance approach.

Population Under Surveillance. The population under surveillance will be youth 0-17 years of age and young adults 18-45 years of age who resided in Colorado at any time in a given calendar year.

Determination of Denominators. Race-bridged post-Censal estimates of the July 1 resident CO population, released yearly by the National Center for Health Statistics (denoted Vintage 20XX release), will be used as the denominators for the Colorado geographic site. Each file contains population estimates for each U.S. county by single year of age, bridged-race, sex, and Hispanic origin. We will categorize race and ethnicity in five categories: Hispanic (any race), ASPI, AIAN, NHB and NHW. Active duty military and institutionalized individuals will be excluded.

Case Ascertainment, Validation, Determination of Diabetes Type and Date of Diagnosis.

Case Ascertainment: The sources of primary case finding for the expanded SEARCH-DiCAYA surveillance system will draw from the well-established network of pediatric and adult endocrinology clinics, community clinics and hospital networks in Colorado established in SEARCH and presented in **Table 1**. Health Data Compass is the data warehouse for University of Colorado Health System integrates the electronic medical record and insurance claims for each of the associated ambulatory and inpatient facilities across the state (which includes the

Barbara Davis Center for Diabetes and the 5 hospitals, 45 outpatient clinics, 14 urgent cares centers and 26 emergency care centers associated with the University of Colorado

Table 1: Colorado Networks for Case Finding	
Outpatient Clinics and Endocrinologists	
Barbara Davis Center (BDC)	Since 1980, the BDC has provided comprehensive care for children, adolescents, and young adults with diabetes, including a dedicated Hispanic/Latino Health diabetes care program and telemedicine to serve patients in remote parts of the state
Pediatric Endocrine Associates (PEA)	6 outpatient clinics along the front range
Rocky Mountain Pediatric Endocrinology	Outpatient offices on the front range
Hospitals (includes inpatient, outpatient clinics and emergency department)	
Valley Wide Health Systems	Located in Alamosa Colorado in the San Luis Valley, serves 15 rural communities
SCL Health (including St. Mary's in Grand Junction, St. Joseph, Lutheran Medical Center in Denver)	5 Hospitals, 64 clinics and 3 ED on the front range, 1 Hospital, 20 clinics and ED on western slope
Health Data Compass (COMPASS) which includes: 1. University of Colorado Hospital (UC Heath) 2. Children's Hospital Colorado (CHOC) 3. Barbara Davis Center (BDC)	Health data warehouse for the University of Colorado Hospital system. Integrates ambulatory and inpatient facilities associated with the University of Colorado Hospital (UC Heath), Barbara Davis Center and Children's Hospital of Colorado (CHOC). Includes 5 hospitals, 45 outpatient clinics (19 outpatient locations for CHOC), 14 urgent cares centers and 26 emergency care centers throughout the state.
Denver Health and Hospital Authority	An integrated, public safety-net institution that has served Denver since 1860 and is currently estimated to care for 1 in 4 adults in Denver. Includes an academic level 1 trauma center, 59 outpatient clinics, 2 emergency departments, 9 federally qualified community health centers, 17 school-based clinics, and health maintenance organization
Centura Health (includes St. Anthony's, North & Central Hospitals in Denver)	Large non-profit healthcare system with facilities across Colorado including: 15 hospitals, 13 affiliate hospitals, 32 outpatient clinics, 23 emergency department and 6 urgent cares centers
Boulder Community Hospital	Located in Boulder Colorado, includes 1 Hospital, 2 outpatient endocrinology practices, 3 emergency departments
Insurance Claims Data	
Colorado All Payer Claims Database (APCD) from the Center for Improving Value in Health Care (CIVHC)	A state-legislated, secure health claims data warehouse containing 100% of Medicaid, >33 commercial health insurance plans. Includes medical and pharmacy claims data.

Health System). As in the SEARCH 1-4 study, key sources of cases for SEARCH-DiCAYA in CO are expected to be the Barbara Davis Center for Childhood Diabetes (BDC) and Pediatric Endocrine Associates (PEA), especially for pediatric cases. However, the Barbara Davis Center for Diabetes diagnoses approximately 100 new patients with T1D annually among adults aged

18-45 years and provides ongoing care to ≥ 3000 young adults with T1D at their center in Aurora, Colorado and at their outreach clinics in Southern Colorado, the Western Slope and Northern CO. Both PEA and the BDC have satellite clinics or telemedicine to serve children in remote areas of the state, allowing wider coverage than would otherwise be possible. Another major source of cases is Children's Hospital of Colorado, which has 19 locations across the front range of Colorado. Denver Health and Hospital Authority (DHHA) will be an important source of case finding for low-income children in Denver who obtain care from the safety-net provider health system. DHHA is estimated to serve 1 in 3 Denver children and during SEARCH, 13.8% of T2D cases were identified at this health care institution. Denver Health is an integrated, public safety-net institution that has served Denver since 1860 with a hospital, 59 outpatient clinics, 2 emergency departments, 9 federally qualified community health centers, 17 school- based clinics, and health maintenance organization. Denver Health is currently estimated to care for 1 in 4 adults in Denver and is expected to be an important source of case finding for T2D. The Colorado All Payers Claims Database is a state-legislated, secure medical claims database containing 100% of Medicaid claims and claims from >33 commercial health insurance plans in CO. All Payers Claims Database is estimated to cover > 60% of the population of Colorado. The state legislation that established CO APCD requires that claims be submitted on a monthly basis, thus APCD is an integral component of an efficient, timely state-wide surveillance effort. Completeness of case ascertainment will continue to be monitored via capture-recapture analyses as described in **Section 3.3.5**.

Numerators: The SEARCH-DiCAYA surveillance system will aim to identify all youth and young adults <46 years of age with incident or prevalent diabetes in Colorado by diabetes type, age, sex, and race/ethnicity on an annual basis starting in 2020 through 2024.

Inclusion criteria for possible/potential cases are as follows:

1. Age eligibility for prevalence: 0-17 years of age (Component A) or 18-45 years of age (Component B) on the last day of the surveillance calendar year (i.e., for surveillance year 2020, age eligibility is 0-17 or 18-45 on 12/31/2020).
2. Age eligibility for incidence: 0-17 years of age (Component A) or 18-45 years of age (Component B) at diagnosis in the index year (i.e., for surveillance year 2020, age eligibility is 0-17 years of age (Component A) or 18-45 years of age (Component B) by 12/31/2020.
3. Residency: The youth or young adult must reside in Colorado at any time during the surveillance calendar year not counting institutionalized residences.
4. Qualifying health care encounter with a diabetes-related ICD-9/ICD-10 encounter diagnosis code, or any diabetes qualifying lab result, or any diabetes qualifying medication prescription or dispensation from 2012 onward. For prevalence, the encounter must have occurred in the 5 years prior to the last day of the surveillance calendar year (i.e., for surveillance year 2020, the diagnosis code must occur between 1/1/2015 and 12/31/2020). For incidence, the estimated onset date must be in the surveillance calendar year.

Exclusion criteria will include cases determined to be active-duty military personnel or institutionalized; if known or recorded based on insurance type associated with active duty military

service. In addition, incidence events among women with gestational diabetes only will not be eligible defined as the delivery date minus 270 days.

Once potential cases are distinguished from cases without a diabetes qualifying event, as outlined above, we will ask our case ascertainment partners to provide a detailed, structured dataset that includes information extracted from electronic medical records including enough PHI necessary to uniquely identify individuals and deduplicate cases across sources, determine the presence of diabetes, determine diabetes type, and date of onset. Given the need to uniquely identify cases, deduplicate cases, and have enough information to determine the presence, type, and date of onset for a diabetes diagnosis, we will request extensive data on people identified as possibly having diabetes including historical encounters dating back to 2012. The rationale for obtaining both diabetes and non-diabetes encounters on the possible diabetes cases identified during the ascertainment window of 2012-present is twofold to insure that we have accurate estimates of diabetes surveillance and that our request fits the context of the minimum necessary PHI because each field is directly tied to either deduplication/uniquely identifying cases, determining the presence of diabetes, determining diabetes type, assessing comorbidities associated with diabetes, and/or assessing the date of diagnosis.

1. Accurate estimates of diabetes surveillance efforts

   A main goal of the project is to evaluate the incidence and prevalence of diabetes by type and evaluate diabetes associated morbidity and mortality including comorbidities, diabetes-related health outcomes and survival, which are currently poorly understood in youth and young adults. Many people with diabetes receive care for diabetes-associated problems like neuropathy, poor wound healing, cardiovascular issues, gastroparesis, and other conditions that may or may not be stated as directly related to their diabetes but factor into gauging the severity of their disease, evaluating how well they are managing their diabetes, or may assist in clarifying diabetes type. Therefore, excluding non-diabetes encounters would lead to an underestimation of health service utilization, identifying co-morbid conditions, and diabetes-related outcomes such as cardiovascular disease, vascular complications, micro and macrovascular kidney damage, and associated end organ damage as well as other conditions complicated by the presence of diabetes, such as COVID-19, which may lead to premature morbidity or mortality. Having non-diabetes encounters are also necessary to understand referral patterns from pediatric to adult centered care, linking youth and young adult diabetes care, which may further explain patterns in early life morbidity and mortality complicated by diabetes. The knowledge gained from non-diabetes encounters will significantly improve the quality of public health significance of people with diabetes in the Colorado.

2. Why non-diabetes encounters meet the minimum data necessary standard for public health surveillance

   Obtaining non-diabetes encounters will not involve collection of additional PHI. This data request meets the definition of the ‘minimum necessary data’ for public health surveillance based on the current practice that protected health information will not be used or disclosed when it is not necessary to satisfy a particular purpose or carry out a function. Each data field requested is tied directly to one of the core pieces of

information needed to produce accurate estimates of diabetes incidence and prevalence by race/ethnicity, age (0-45), sex, and diabetes type. The confidential information requested represents the minimum information necessary for the Colorado School of Public Health to perform the duties described in the approved protocol and only the minimum necessary individuals shall have access to the confidential information in order to perform such work. The minimum necessary standard requires covered entities to evaluate their practices and enhance safeguards as needed to limit unnecessary or inappropriate access to and disclosure of protected health information. The data sharing agreements and privacy protection practices for this surveillance project have been developed by a team of clinical and public health partners in Colorado. Our team represents multiple health care and public health systems in Colorado and our resulting protocol reflects the disclosure and privacy protection standards requested by each agency.

The specific fields requested and detailed justifications for the requested fields are included below.

Table 2: Detailed Electronic Health Record Data Fields Requested and Associated Uses for Potential Diabetes Cases Identified through Inclusion Criteria

Electronic Health Record data fields requested	**Reason**
Medical Record Number	Unique identifier to link across tables, may be random identifier
Full Name (Last Name, First Name, MI)	Needed for deduplication purposes to uniquely identify each case across sources
Date of Birth	Needed for deduplication purposes to uniquely identify each case across sources, to determine age 0-45 eligibility in each year, and to classify age at diagnosis
Sex	Needed for deduplication purposes to uniquely identify each case across sources and to classify cases by sex
Race and ethnicity (if known)	Needed to classify cases by race/ethnicity
Address history (including Zip Code and County; and dates of use)	Needed for deduplication purposes to uniquely identify each case across sources and to determine residence in Colorado eligibility
Membership in U.S. military or part of an institutionalized population (i.e. inmate at a prison/jail/juvenile lockup, resident at a group, foster home, long term care facility or halfway house)	Needed, if known, to exclude cases not included in denominators as part of military or institutionalized populations otherwise excluded from Census figures.

All ICD-9-CM and ICD-10-CM diagnostic codes from 2012 to 2020 *Excluding mental health admissions/care, drug or rehabilitation associated care, treatment or testing for sexually transmitted infections, or any visit associated with child abuse or domestic violence unlikely to be related to diabetes.*	All diagnosis codes are needed from 2012 onward to assist with determinations of the presence of diabetes, distinguish diabetes type, and date of onset. All diabetes codes are needed to determine diabetes type and date of onset. Non-diabetes diagnoses are needed to better contextualize health information that may preclude the determination of a diabetes diagnosis such as co-occurring conditions that could alter diabetes type classification such as diabetes diagnosed secondary to cystic fibrosis, gestational diabetes, or autoantibody positive T2D that otherwise might be mistyped as T1D. Further idiosyncratic elevations in glucose levels, unrelated to diabetes, may be caused by temporary high dose steroid use, trauma, or immediately following surgery. Supplementary diagnoses provide necessary context to correctly interpret data that could otherwise lead to a misclassification of diabetes type and/or presence of diabetes.
Admission or Visit Date(s) for **all encounters and procedures** (including inpatient, outpatient, or emergency/urgent care visits including any encounters or procedures for SARS Cov2/COVID-19) from 2012 to 2020 *Excluding mental health admissions/care, drug or rehabilitation associated care, treatment or testing for sexually transmitted infections, or any visit associated with child abuse or domestic violence unlikely to be related to diabetes.*	All visit dates are needed from 2012 onward to assist with determinations of the presence of diabetes, distinguish diabetes type, and date of onset. All dates for diabetes codes are needed to determine date of onset and contextualize diabetes type when more than one diabetes type code is present, and assessment of diabetes type may have changed over time. Dates associated with non-diabetes diagnoses are needed to better contextualize co-occurring health information that may preclude the determination of a diabetes diagnosis such as concurrent use of high dose corticosteroids to treat COVID-19, or a co-occurring pregnancy which would change a diagnosis to gestational diabetes that is specific to pregnancy, but possibly initially miscoded as T2D unless the diabetes predated the pregnancy.

All diabetes and COVID related medications dispensed and prescribed from 2012 to 2020 (see attached list of medication classes)	All prescribed medications are needed from 2012 onward to assist with determinations of the presence of diabetes, distinguish diabetes type, and date of onset. Non-diabetes specific medications are needed to better contextualize health information that may preclude the determination of a diabetes diagnosis such as concurrent high dose steroids used in treatment for COVID-19, which can temporarily elevate glucose levels causing false positives for temporally associated laboratory testing. Similarly, knowing the reasons for prescribed medications can be useful as diabetes associated medications are prescribed for other reasons. For example, Metformin, commonly prescribed for T2D, is also prescribed for PCOS, treatment of obesity, or preventatively for pre-diabetes in people at risk for T2D, but not clinically diagnosed as case of diabetes. Similarly, insulin, most often associated with T1D treatment, is sometimes used in treatment of trauma or other emergent contexts unrelated to diabetes.
All diabetes and COVID related laboratory results from 2012 to 2020 (see attached list)	All requested laboratory results are needed from 2012 onward to assist with determinations of the presence of diabetes, distinguish diabetes type, and date of onset. These values, in conjunction with co-occurring health information and diagnoses, can assist with the correct classification of the presence or absence of diabetes, diabetic ketoacidosis, and rule out temporary hyperglycemia from corticosteroid use, trauma, or other contexts not caused by diabetes.
All clinical/progress notes for encounters included in query from 2012-2020 *Excluding mental health admissions/care, drug or rehabilitation associated care, treatment or testing for sexually transmitted infections, or any visit associated with child abuse or domestic violence unlikely to be related to diabetes.*	All progress notes from included encounters are needed from 2012 onward because these notes often include supporting information that may be useful in determining the presence of diabetes, determining diagnosed diabetes type, and/or date of onset with supplemental free text information not otherwise included in the fields requested. For example, these notes will sometimes include historical information such as, "Patient X is initiating care today for diabetes diagnosed at age 32 in 2011 when she was living in Florida." This contextual information would assist in the correct classification of incident year and allow us to include the case as prevalent in the year(s) of interest, but not incident because in the example above the person lived outside of Colorado in the year they were diagnosed.

All problem list diagnoses and conditions from 2012 to 2020 *Excluding mental health admissions/care, drug or rehabilitation associated care, treatment or testing for sexually transmitted infections, or any visit associated with child abuse or domestic violence unlikely to be related to diabetes.*	All problem list diagnoses are needed from 2012 onward to assist with determinations of the presence of diabetes, distinguish diabetes type, and date of onset. Diabetes codes are needed to determine diabetes type and date of onset, however non-diabetes diagnoses are also needed to better contextualize health information that may alter the determination of a diabetes diagnosis such as co-occurring conditions that could change diabetes type classification such as diabetes diagnosed secondary to cystic fibrosis, gestational diabetes occurring only in the context of pregnancy, or autoantibody positive T2D that otherwise might be mistyped as T1D. Further idiosyncratic elevations in glucose levels, unrelated to diabetes, may be caused by temporary high dose steroid use, trauma, or immediately following surgery. Supplementary diagnoses provide necessary context to correctly interpret data that could otherwise lead to a misclassification.
Records of vitals and clinical observations (e.g.- height, weight, blood pressure, temperature, respiration rate, etc.) from 2012 to 2020 (see attached list)	The requested clinical observations from 2012 onward are necessary to help assist with determining diabetes type and/or gauge the presence of and severity of DKA episodes.
Date of diabetes diagnosis (if known)	This field is useful in determining the date of diagnosis, if available.
Date of death (if applicable and known)	Date of death is useful in determining mortality status as part of our secondary aims.
Cause of death (if applicable and known)	Cause of death is useful in determining diabetes associated vs. non-diabetes associated mortality.

Algorithm to Identify potential diabetes cases: In March of 2021, and annually thereafter, each data source will be queried to identify "potential" diabetes cases occurring from 2012 through the last calendar year (the first query in March of 2021 will be for surveillance calendar years 2012-2020, in March 2022 will be for surveillance calendar years 2013-2021, etc.) using a "Wide Net" algorithm (**Table 2**). Each data partner will be asked to provide a structured dataset containing all diabetes relatedencounters for each identified possible case occurring from 2012 to the current surveillance year (for example in surveillance year 2020, the dataset would include all diabetes-related encounters occurring between 1/1/2012-12/31/2020). The variables for the healthcare encounters will be grouped into the following domains: diagnostic codes, encounter dates, laboratory measurements, clinical characteristics, demographics, medications, personal identifying information. If the data partner is uncomfortable exchanging personal identifying information, analytic support will be provided to implement a hashing procedure to protect individual privacy while allowing the SEARCH-DiCAYA team to link patient encounter data with both the current SEARCH Diabetes Registry and across data sources.

Table 3. First step of SEARCH Wide Net Algorithm to Identify Potential Diabetes Cases.
Individuals are considered to have possible diabetes if they meet any of the following: criteria
• Hemoglobin A1c >=6.5% (>=47.5 mmol/mol)
OR
• Fasting plasma glucose >=126mg/dl (>=7.0 mmol/L)
OR
• Random plasma glucose >= 200 mg/dl (>=11.1 mmol/L)
OR
• At least one of the following diabetes-related ICD-10 encounter diagnoses codes: E08, E09, E10, E11, or E13.
OR
• A prescription for or administration of one or more medications residing in one of the following antidiabetic medication classes*:
• Metformin
• Glyburide/glibenclamide
• Glimepiride
• Glipizide
• Gliclazide
• Glyclopyramide
• Gliquidone
• Chlorpropamide
• Tolazamide
• Tolbutamide
• Acarbose
• Miglitol
• Voglibose
• Alogliptin
• Linagliptin
• Saxagliptin
• Sitagliptin
• Vildagliptin
• Nateglinide
• Repaglinide
• Canagliflozin
• Dapagliflozin
• Empagliflozin
• Ertugliflozin
• Rosiglitazone
• Pioglitazone
• Troglitazone
• Pramlintide
• Albiglutide
• Dulaglutide
• Exenatide
• Liraglutide

• Lixisenatide
• Semaglutide
• Insulin
• Glucagon

The second step of our case ascertainment process involves obtaining the detailed EHR data on cases identified as having a diabetes qualifying event, as outlined above, to confirm the presence of diabetes, determine diabetes type, and determine date of onset.

Record linkage across data sources: Incremental record linkage will be used to reconcile identify and link potential duplicate cases identified across data sources and within the current SEARCH Registry. It is anticipated that a significant proportion the youth and many young adults born from 1982 onward identified by the Wide Net algorithm as possible cases will have previously been identified by SEARCH and contained in the archival SEARCH Registry (2001-2020). Linkage to the SEARCH Registry will be useful as it will help identify true incidence cases from prevalent cases that may be new to a healthcare system but not true incident cases if diagnosed elsewhere or in a different year. The incremental record linkage strategy will be implemented using a hybrid deterministic and probabilistic approach, designed to be both efficient and increase the accuracy of linkage for cases with missing data in key personal health identifiers as well as to deduplicate across sources. Newly identified cases (not determined to have a previous match in the SEARCH Registry) will be assigned a unique patient identifier and evaluated for additional diabetes-related encounters across data sources. Encounters belonging to the same individual will be combined to create a timeline of diabetes-related healthcare encounters over a 5-year observation period preceding the index surveillance year designed to assist with determining date of onset and rule out alternative reasons for uses of diabetes-associated medications or elevated laboratory values associated with hyperglycemia. The timeline will allow for more accurate identification of both date of diagnosis and diabetes type by the ICD-10 based algorithms which rely upon the frequency and timing of diabetes diagnostic codes. The incremental aspect of record linkage applies to the design and improvement the unique patient identifier for each subsequent year of surveillance, thus once a patient is identified as an incident case from data source A in year 1, the internally generated unique identifier will flag that individual as a potential prevalent case in subsequent years of ascertainment. The application of the "Wide Net" algorithm as a first step to identify potential cases and then the creation a timeline of diabetes-related care is anticipated to increase sensitivity and PPV of ascertainment of T2D cases may have intermittent care or cross healthcare systems.

Table 4. Performance of a rule-based algorithm using ICD-10 codes for determining presence of diabetes using the SEARCH ascertained provider type as the gold standard - Colorado.

		Rule-based algorithm of ≥2 ICD diabetes codes (n=2680)
Diabetes n = 5,308	*Se*	0.997
	Sp	0.984
	PPV	0.991
	NPV	0.990

Refined algorithm to determine probable diabetes cases: A simple rule-based algorithm of two or

more diabetes-related ICD codes will be applied to the integrated diabetes-related healthcare encounter timelines of each individual identified as a potential diabetes case by the "Wide Net" algorithm to ascertain probable diabetes cases. Table 4 demonstrates the high performance of the rule-based algorithm of ≥2 ICD diabetes codes to identify "true" SEARCH ascertained diabetes cases in Colorado.

Algorithm to determine of diabetes type: Among youth ascertained to be probable diabetes cases, the most frequently occurring diabetes specific ICD-10 code will be used to classify individuals (based on their diabetes-related healthcare encounter timelines) into five mutually exclusive categories of probable diabetes type: not diabetes, only or primarily T1D, only or primarily T2D, type undetermined (if the ratio of T1D and T2D ICD-10 codes is 1.0) and other diabetes types.

Case validation via targeted chart review: To improve the overall performance of the ICD-10 based algorithms to determine diabetes type study staff will perform a targeted chart review of a subset of newly identified possible diabetes cases identified by the ICD-10 to confirm the presence of diabetes, confirm diabetes type, and/or to clarify date of onset. Validation efforts will occur both for case versus non case classifications as well as validations of type and date of onset for confirmed cases. Study staff will abstract information from the medical record for the purposes of ascertaining diabetes type and date on onset from the period of the first diabetes ICD-10 code to 6 months after this date. Information reviewed in the chart will include: 1) clinical notes, 2) results of diabetes autoantibody measurement (GAD65/GAA, IA2/ICA512, ICA, IAA, and ZnT8), 3) height and 4) weight (closest to diagnosis), 5) whether the participant ever used insulin, 6) whether insulin was discontinued, 7) presence of acanthosis nigricans, and 8) whether DKA was noted (with dates, bicarbonate, pH, and glucose values). A structured data collection form will be completed by the data abstractor with fields for determined diabetes type and onset date.

Determination of prevalent and incident cases: In SEARCH-DiCAYA, we will distinguish between incident and prevalent cases using two principal methods: 1) matching to the existing SEARCH Registry and 2) use of an EHR- algorithm that predicts date of diagnosis based on the calendar year of first diabetes ICD-9/-10 code supplemented by data from the targeted chart review of cases where presence of diabetes, diabetes type, or date of diagnosis is unclear. New onset cases will be classified as both incident and prevalent and preexisting diagnosed cases, both those identified previously in Colorado and those previously identified elsewhere, will be classified as prevalent.

Ascertainment of core variables: A minimum amount of demographic and clinical information is needed for all cases in order to calculate population-based incidence rates and prevalence of diabetes mellitus by age, sex, diabetes type and race/ethnicity. Each case finding data source will be asked to provide a structured dataset containing encounters necessary to contextualize diabetes presence and type, demographic and limited clinical information for each identified case identified by the "Wide Net" algorithm. Demographic data will include: 1) full name, 2.)date of birth, 3) sex, 4) race/ethnicity, 5) residential addresses. EHR data will include the following information on requested healthcare encounters dating back to 2012: 1) diagnosis date (if available) 2) all diagnostic codes, 3) all encounter dates, 4) diabetes related medications and

reason for prescription, 5) all relevant laboratory results, 6) all progress notes, 7) mortality status and cause of death, if known. Other clinical information requested will include: 1) vitals and clinical observations (e.g.,-height and weight), 2) diabetes autoantibody testing results, and 3) presence of acanthosis nigricans.

Ascertainment of DKA at diagnosis: As part of the structured dataset requested from each case-finding data source we will request the following information on all cases identified by the "Wide Net" algorithm to assess presence of DKA among cases determined to be newly onset. ICD-10 codes to the second decimal place, encounter dates for all diagnoses, laboratory values related to DKA assessment (bicarbonate, pH, and glucose values). DKA encounters will be included as part of all encounters outlined in the second part of the two-step case ascertainment process.

Determination of Eligibility**:** Eligibility for prevalent diabetes will include a diabetes-related ICD-9 or 10 encounter diagnosis code in the 5 years prior to the last day of the surveillance calendar year (i.e., for surveillance year 2020, the diagnosis code must occur between 1/1/2015 and 12/31/2020) in conjunction with age eligibility (< 45 years of age on the last day of the surveillance calendar year) and a residential address in Colorado at any time in the surveillance calendar year. For incidence, the estimated onset date must be in the surveillance calendar year in conjunction with age eligibility at the time of diagnosis and a residential address in Colorado at any time in the year of diagnosis. Youth and young adults who are determined to be active duty military personnel or institutionalized in the surveillance calendar year/ year of diagnosis will not be eligible.

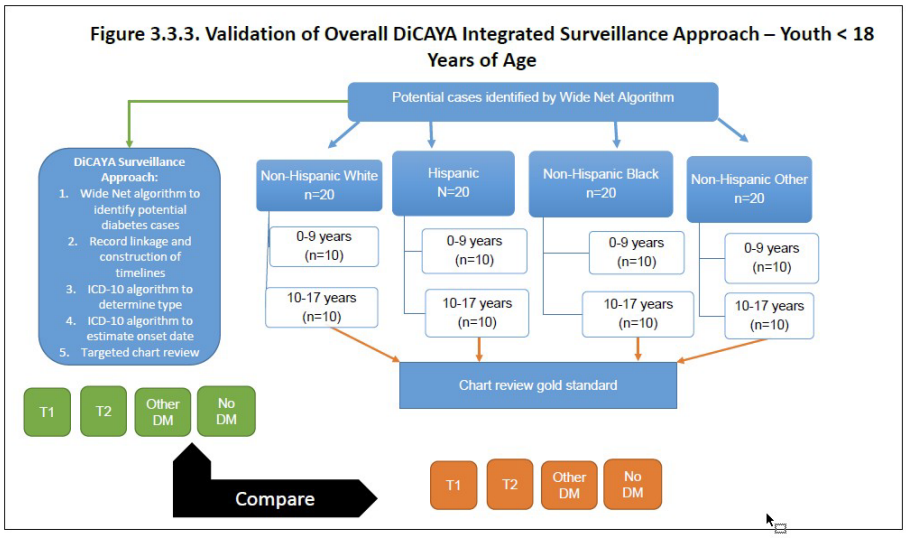


Figure 3.3.3. Validation of Overall DiCAYA Integrated Surveillance Approach – Youth < 18 Years of Age

Validation of Overall SEARCH-DiCAYA Integrated Surveillance Approach. To address our third specific aim we will implement a case validation protocol with the goal of determining the sensitivity, specificity and positive predictive value of the SEARCH-DiCAYA surveillance approach, both overall and by diabetes type compared to a gold standard of chart review. **Figures 3.3.3**. and **3.3.4.** Illustrate how we will compare diabetes type as ascertained by the SEARCH-DiCAYA surveillance approach to diabetes type ascertained from chart review among a random sample of 520 potential diabetes cases identified by the "Wide Net" algorithm, by component stratified by race/ethnicity and age group for components A and B, respectively. Chart review will be conducted by trained staff and supervised by a physician-scientist (Dr. Dabelea). The chart abstractor(s) will compete a standardized medical record abstraction form based on data in diagnostic codes, encounter dates, laboratory measurements (including HbA1C and antibody testing), clinical characteristics, demographics, medications, and physician note fields. We will work with other funded sites and CDC to determine the ultimate criteria for the gold standard determination of diabetes type in mutually exclusive categories of T1D, T2D, other DM and no DM. **Table 5** demonstrates how we will calculate and interpret the measurements of accuracy and validity of the SEARCH-DiCAYA surveillance approach, using the example of T1D.

Figure 3.3.4. Validation of Overall DiCAYA Integrated Surveillance Approach – Young Adults

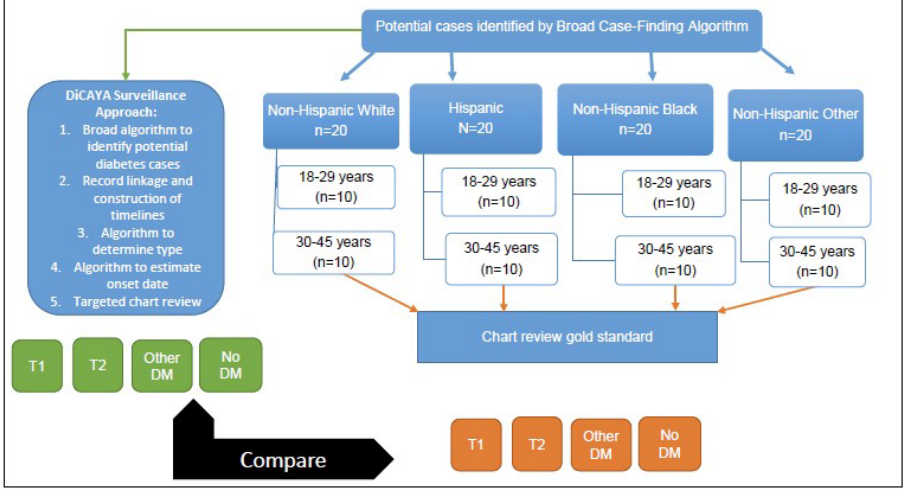


Table 5 Determination of performance of the SEARCH-DiCAYA surveillance approach to identify T1D.			
		Gold Standard Chart Review Determination of T1D from Potential Cases Identified by Wide Net Algorithm	
SEARCH-DiCAYA Surveillance Approach		+	-
	+	TP= T1D identified by DiCAYA that agree with chart review	FP=T1D cases identified by DiCAYA that don't have DM per chart review
	-	FN= T1D cases missed by DiCAYA	TN= potential cases identified by Wide Net that are ascertained by DiCAYA and chart review to not have T1D
Sensitivity= the likelihood that a T1D case is identified by the DiCAYA surveillance approach.			
Specificity=the likelihood that a youth identified by the "Wide Net" algorithm as a potential diabetes case, who does not have T1D, is correctly determined so, by the DiCAYA surveillance approach.			
PPV=the likelihood that a T1D cases identified by DiCAYA surveillance approach is truly a T1D case			
NPV=the likelihood that a youth identified by the "Wide Net" algorithm as a potential diabetes case and determined by DiCAYA surveillance approach to not have T1D, truly does not have T1D			

Investigating Potential Bias Introduced by Less than Complete Case Ascertainment.

Surveillance systems developed with EHR or claims data often fail to capture all eligible cases due to limitations in case ascertainment methods. In order to increase the accuracy of the SEARCH-DiCAYA incidence and prevalence estimates, it will be crucial to identify individuals from the target population not captured by our primary case-finding data sources. Failing to account for these 'missed cases' would underestimate the true population proportion of youth and young adults with diabetes and consequently, underestimate the true prevalence and incidence.(50, 54) The magnitude of bias may also be related to individual patient characteristics and is expected to differ by diabetes type. Youth and young adults with diabetes that experience gaps in routine medical follow-up, or those who have a change in health insurance would be less likely to be identified as a case an EHR-based surveillance system. We will use a two-mode capture-recapture approach used by SEARCH (52), which has provided the most consistent estimates of completeness across funded sites and over time. The approach classifies if a case was cared for in an outpatient setting "mode" or an inpatient/hospital setting "mode".

Adjusted log-linear methods will be used to estimate the total size of the population of youth with diabetes aged 0-17 and adults 18-45 in the target population. We anticipate improvement of capture-recapture estimates in the SEARCH-DiCAYA surveillance approach afforded by the hybrid deterministic and probabilistic incremental record linkage across primary case-finding data sources. Capture-recapture techniques are highly dependent on record linkage performance as the number of individuals missed by a surveillance system is estimated using the overlap between "modes." Imperfect record linkage may result in misclassification of unique individuals across modes which may be especially important for T2D cases who seek healthcare across settings and may have incomplete or erroneous values for key PHI field that prevented previous matches. The population of individuals with diabetes estimated from this approach will be used to calculate the capture-recapture adjusted prevalence and incidence estimates.

Database Management and Quality Control. A standardized data request will be used to ascertain potential diabetes cases from primary case-finding data sources using the "Wide Net" criteria case definition. Datasets from each of our partners will be transmitted to the Lifecourse Epidemiology of Adiposity and Diabetes (LEAD) Center using SSL (Secure Sockets Layer) certification to ensure a secure file transfer on an encrypted connection, over encrypted email, or by mail on encrypted CDs. Encounter-level data will then be obtained and formatted using uniform data fields, internal formatting constraints, and specified allowable values or ranges. Data elements received from each data source will be collated and data quality will be examined within and across sources to extract summary variable values using predefined criteria. Patient level surveillance data will be stored at the LEAD Center at the University of Colorado in a password-protected HIPAA-compliant database housed on a secure internal server with restricted access for approved study personnel only. In compliance with the HIPAA Security Rule, we previously conducted a risk assessment to evaluate the inherent risk of storing clear-text PHI data at the University of Colorado and developed a Security Plan addressing threats and vulnerabilities which is updated on a yearly basis. De-duplicated, limited individual and summary-level data will be generated and submitted to the DiCAYA coordinating center at New York University (NYU) via a secure network on a periodic basis as prescribed by CDC as described in the data use agreement between UCD and NYU.

The data elements we have been requested by NYU to share in a limited data set would include:

1. Patient month and year of birth: to define patient age for each measurement year (necessary for component age inclusion criteria)
2. Patient sex: to validate diabetes incidence and prevalence phenotypes estimates by sex
3. Patient race/ethnicity: to validate diabetes incidence and prevalence phenotypes estimates by race/ethnicity
4. Patient address county: to provide diabetes incidence and prevalence estimates by county and for bias adjustment methods
5. Patient address state: to provide diabetes incidence and prevalence estimates by state and for bias adjustment methods
6. Patient address 5-digit zip code: primarily for bias adjustment methods and potentially to provide diabetes incidence and prevalence estimates by zip code if necessary
7. Encounter type (ambulatory, emergency department, inpatient, etc.): For visits with a diabetes diagnosis code, diabetes-indicated medication, or laboratory value indicating diabetes within 3 years of the incidence year, to potentially exclude diagnosis from certain encounter types and to confirm presumed diabetes and diabetes type, which rely on different criteria for inpatient versus outpatient diagnosis codes.
8. Admission/encounter date (month, day, and year): For visits with a diabetes diagnosis code, diabetes-indicated medication, or laboratory value indicating diabetes within 3 years of the incidence year to define year of diabetes incidence for calculation of incidence rates
9. Provider specialty: For visits with a diabetes diagnosis code, diabetes-indicated medication, or laboratory value indicating diabetes within 3 years of the incidence year to resolve conflicts in diabetes type classification based on prioritizing endocrinologist-assigned diabetes type.
10. Diagnosis type (ICD-9-CM, ICD-10-CM, etc.): For visits with a diabetes diagnosis code within 3 years of the incidence year to define diabetes type, which relies on distinguishing type 1, type 2, and other diabetes type diagnosis codes
11. Diagnosis code: For visits with a diabetes diagnosis code within 3 years of the incidence year to define diabetes type and exclusionary conditions.
12. Diagnosis source (admitting, discharge, etc.): For visits with a diabetes diagnosis code within 3 years of the incidence year to subset to reliable diagnoses (e.g., final or discharge diagnoses) when classifying diabetes type.
13. Diagnosis date (month and year): For visits with a diabetes diagnosis code within 3 years of the incidence year to define year of diabetes incidence for calculation of incidence rates.
14. Laboratory result specimen source: For laboratory values indicating diabetes within 3 years of the incidence year to subset to blood specimen sources to define diabetes-related glucose tests for classifying diabetes type.
15. Laboratory result LOINC code: For laboratory values indicating diabetes within 3 years of the incidence year to define diabetes-related labs (C-peptide, diabetes autoantibody, A1c, random blood glucose, fasting blood glucose) for classifying diabetes type.
16. Laboratory result order date: For laboratory values indicating diabetes within 3 years of the incidence year to define diabetes type, sub-setting to labs prior to the end of the measurement year when specimen collection date is unavailable.

17. Laboratory result date of specimen collection: For laboratory values indicating diabetes within 3 years of the incidence year to define diabetes type, sub-setting to labs prior to the end of the measurement year.
18. Qualitative value of laboratory result: For laboratory values indicating diabetes within 3 years of the incidence year to define negative C-peptide or positive diabetes autoantibody for classifying diabetes type.
19. Numeric value of laboratory result: For laboratory values indicating diabetes within 3 years of the incidence year to define elevated A1c, fasting blood glucose, random blood glucose for classifying diabetes type.
20. Unit for numeric value of laboratory result: For laboratory values indicating diabetes within 3 years of the incidence year to define elevated A1c, fasting blood glucose, random blood glucose for classifying diabetes type.
21. Abnormal laboratory result indicator: For laboratory values indicating diabetes within 3 years of the incidence year to define elevated A1c, fasting blood glucose, random blood glucose for classifying diabetes type.
22. Laboratory result modifier (equal, greater than or equal, etc.): For laboratory values indicating diabetes within 3 years of the incidence year to define elevated A1c, fasting blood glucose, random blood glucose for classifying diabetes type.
23. Raw laboratory result name: For laboratory values indicating diabetes within 3 years of the incidence year to define diabetes-related labs when LOINC codes are unavailable (C-peptide, diabetes autoantibody, A1c, random blood glucose, fasting blood glucose) for classifying diabetes type.
24. Prescription order date: For diabetes-indicated medications prescribed within 3 years of the incidence year to define diabetes type, sub-setting to medications prior to the end of the measurement year when start date is unavailable.
25. Prescription start date: For diabetes-indicated medications prescribed within 3 years of the incidence year to define diabetes type, sub-setting to medications prior to the end of the measurement year.
26. Raw prescription medication name: For diabetes-indicated medications prescribed within 3 years of the incidence year to define diabetes-related prescriptions when RxNorm codes are unavailable for classifying diabetes type.
27. Prescription RxNorm Concept Unique Identifier: For diabetes-indicated medications prescribed within 3 years of the incidence year to define diabetes-related prescriptions for classifying diabetes type.
28. Medication administration start date: For diabetes-indicated medications prescribed within 3 years of the incidence year to define diabetes type, sub-setting to medication administrations prior to the end of the measurement year.
29. Medication administration code type (NDC, RxNorm, etc.): For diabetes-indicated medications prescribed within 3 years of the incidence year to define diabetes-related medication administrations for classifying diabetes.
30. Medication administration code: For diabetes-indicated medications prescribed within 3 years of the incidence year to define diabetes-related medication administrations for classifying diabetes type.
31. Raw medication administration medication name: For diabetes-indicated medications prescribed within 3 years of the incidence year to define diabetes-related medication administrations when administration codes are unavailable for classifying diabetes type.

IV. DATA ANALYSIS PLAN

Computation of Annual Prevalence Estimates – Aim 1. Prevalence estimates will be derived by sex-, age- and race/ethnicity groups within each diabetes type. The prevalence will be expressed per 1,000 youth aged 0-17 and young adults aged 18-45 in Colorado and 95% CI will be calculated using a skew-corrected inverted score test assuming a binomial distribution.(55) Trends in prevalence will be assessed by comparing the SEARCH prevalence estimates in 2001, 2009 and 2017 to the estimates derived from DiCAYA using Poisson regression models. Standard errors associated with estimated change in prevalence between any two time points will be computed using a 2-sided skew-corrected inverted score tests for binomial distribution. Standard error for the trends in prevalence estimates will be derived from the Poisson regression model. This model will also be used to generate adjusted prevalence where adjustment will be made for race/ethnicity, age and sex. Assuming a linear relationship between prevalence estimates over time, the detectable rate of change was estimated using the approach proposed by Nam(56), which can be seen as a generalization of Cochran-Armitage's trend test for linear trends in proportions. **Table 6** shows the detectable rate of change in prevalence by diabetes type in Colorado between 2001 and 2024 assuming a type 1 error rate of 5%, power levels of 90% and 80%.

Table 6: Detectable difference in prevalence of diabetes by type

Diabetes type	Prevalence per 1,000					Δ 2001-2020	Δ 2001-2025	Detectable Δ(%)	
								Power = 90%	Power = 80%
Year	2001	2009	2017	2020	2025				
T1D	1.55	2.07	2.53	2.71	2.91	74.8%	87.7%	3.2	2.6
T2D	0.07	0.13	0.13	0.14	0.15	100.0%	114.3%	6.1	5.7
*Prevalence estimates assume a linear relationship between prevalence rate and time									

Assuming that the annual linear increase in the prevalence of T1D observed in Colorado between 2001 and 2017 remains at 0.6125 per 1,000 annually, the prevalence in 2020 is estimated at 2.71 per 1,000 and in 2024 at 2.91 per 1,000 youth. Assuming the annual linear increase in the prevalence of T2D observed in Colorado between 2001-2017 is 0.00375 per 1,000 youth, the estimated prevalence in 2020 is 0.14 and in 2024 is 0.15 per 1,000 youth. The comparison between these prevalence estimates and the detectable level of changes suggests that we are well-powered to detect changes in prevalence for each diabetes type. For example, we will have at least 90% power to detect an absolute % change in the prevalence of T1D in youth of 3.2% and 6.1% among youth with T2D between 2001-2020.

Computation of Annual Incidence Estimates – Aim 2. A similar approach will be taken to estimate the incidence rates of diabetes by type, race/ethnicity, sex and age. Incidence rates will be estimated as the number of unique newly diagnosed cases divided by the total number of individuals age 0-17 and 18-45 in Colorado, expressed per 100,000 individuals, and will be stratified by age, sex, and race-ethnicity. Trends in incidence will be estimated using generalized autoregressive moving average (GARMA)

Table 7: Detectable rate of change in incidence by diabetes type in youth 0-17

T1D		T2D	
Power=90%	Power=80%	Power=90%	Power=80%
1.3	1.1	0.85	0.74

models described by Mayer-Davis (3). We will estimate trends in incidence from each surveillance year since 2002. Trends will be adjusted for age, sex, and race or ethnic group and unadjusted trends will be estimated with a negative binomial distribution with logarithm link. The detectable effect size for each power level was estimated using the same approach described in section 5. SEARCH-DiCAYA in Colorado will have 90% power to detect a difference in the incidence of T1D between 2002 and 2020 of 1.3 per 100,000 youth. We will have 90% power to detect a difference in the incidence of T2D of 0.85 per 100,000 youth between 2002 and 2020.

Prevalence of DKA at Diagnosis Among Newly Diagnosed T1 and T2 Diabetes Cases: Between 2010 to 2016, SEARCH observed a 2% annual increase in DKA at diagnosis among T1D cases, with an absolute change in DKA prevalence of 5.3%. The prevalence of DKA among incident T1D and T2D cases will be obtained from primary case finding data sources, using prior SEARCH definitions based on laboratory values (PH and bicarbonate levels) and/or ICD-10 diagnostic coding E10. The prevalence of DKA at onset of diabetes will be estimated by diabetes type, age, sex, and race/ethnicity groups with logistic regression and with absolute differences between subgroups calculated from age-adjusted incidence estimates. Assuming a significance level of 0.05, we would have 80% power to detect an absolute change of 0.22% (from 30.3% to 30.1% for example) in the prevalence of DKA among T1D cases, and an absolute change of 0.25 (from 7.2% to 6.95%) in T2D cases between 2002 and 2020.

Table 8: Detectable rate of change in incidence by diabetes type in adults 18-45

T1D		T2D	
Power=90%	Power=80%	Power=90%	Power=80%
3.2	2.7	2.0	1.7

Evaluation of SEARCH-DiCAYA Surveillance Methods – Aim 3.

Adjusting Results using Capture-Recapture Methods. Capture-Recapture adjusted estimates of prevalence and incidence will be obtained from the two-mode approach used by SEARCH by classifying each case in terms of identification in an outpatient setting "mode" or an inpatient/hospital setting "mode". Adjusted log-linear methods will be used to estimate the total size of the population of youth with diabetes aged 0-17 and 18-45 in the target population. Analysis for the second analytical approach will entail identifying dependency between each case-finding data source in our study by estimating capture-recapture odds ratios.(51) Potential sources of heterogeneous capture probability will be evaluated for age, sex, race/ethnicity and insurance status by testing for an association across case-identifying data sources. Base and additional models will be evaluated by including terms for the data source and two-way interactions between data source and the factors that are found to be significantly associated with heterogeneous capture probability. Interactions between each data source combination found to be dependent will also be included. Final model selection will be determined based on the model with the lowest Akaike Information Criteria (AIC).(51) From both approaches, the "true" population of individuals with diabetes will be estimated from the final models and the capture-recapture adjusted estimate will be obtained by dividing the estimated number of youth and young adults with diabetes by the population aged 0-17 and aged 18-45 years in Colorado.

V. DATA MANAGEMENT PLAN (DMP)

The data that will be produced as part of this project include:

- Local SEARCH-DiCAYA Diabetes Surveillance Database: an encounter and summary-level

database of individuals with diabetes aged 0-17 and 18-45 in Colorado during each year of case ascertainment. The minimum amount of PHI necessary to uniquely identify cases, determine values of interest (e.g.- presence of diabetes, determine diabetes type, and date of diagnosis) and record necessary demographics will be locally maintained in a secure, password protected database.

- Clear-text PHI needed for record linkage and de-duplication of diabetes cases within and across data sources
- Limited SEARCH-DiCAYA Diabetes Surveillance Database: The limited SEARCH-DiCAYA Diabetes Surveillance Database will include limited information necessary to be made available to the NYU coordinating center, the study sponsors, and/or researchers in need of access to de-identified data compiled from the limited data sets shared with NYU. Both the local SEARCH-DiCAYA Diabetes Surveillance Database, with identifiers, and the limited SEARCH-DiCAYA Diabetes Surveillance Database would be password protected with user-based role accessibility and housed on the University of Colorado's secure Isilon server that is HIPAA compliant.

Data will be collected according to strict data standards to ensure reliability and reproducibility following both local and national DiCAYA protocols. In addition, a complete description of the methods of data collection, data dictionaries, and potential limitations of the data will be documented. The local SEARCH-DiCAYA Diabetes Surveillance Database will be developed by 12/31/2021. The clear-text PHI data will be received from primary case-finding data sources annually by November 30 starting in 2021. The University of Colorado will house the local SEARCH-DiCAYA Diabetes Surveillance Database and the limited diabetes surveillance data repository in a password-protected database on a secure internal server at the LEAD Center. The computers of the University of Colorado research team are on a segregated network utilized specifically for data that falls under the security rules of HIPAA. The University of Colorado uses network segregation as means of data separation as per HIPAA requirements. This network is disconnected from the standard, public University network using firewalling and secure routers. Additionally, the University of Colorado uses a system of "access control" for certain folders that are located within the HIPAA network that will house the local SEARCH-DiCAYA Diabetes Surveillance Database and the SEARCH-DiCAYA data repository on the secure internal server. The folders are access restricted. Permissions are organized and granted by the IT department. Access to individual files and folders are assigned unique permissions stored in the Active Directory. The University of Colorado's IT team will authorize members of the research team to have access to a specific folder housing the project data. Access to this folder is restricted to the research team, and users require passwords to access this folder. The passwords are a nonsensical combination of numbers and letters, changed on a regular schedule, never repeated, and stored away from the computer.

For the clear-text PHI data, additional security measures will be implemented. Briefly, we will create a virtual machine that is not an internet-facing server using a full-blown jump box desktop. For all technical and security standpoints, it will meet every detail specified in the Security Plan. The Privacy Rule permits assigning to, and retaining with, the health information a code or other means of record identification if that code is not derived from or related to information about the individual and could not be translated to identify the individual. The

Security Rule operationalizes the protections afforded in the Privacy Rule by establishing standards for addressing the technical and non-technical safeguards that organizations must have in place to protect the privacy of individuals' PHI. The project team will implement and oversee policies, procedures, and technologies that are appropriate for their identified risks to PHI. The University of Colorado will remove the all PHI, or any other means of record identification, prior to uploading the data into the SEARCH-DiCAYA surveillance system so that there will be no method to re-identify the data. Only the limited data will be transferred to the SEARCH-DiCAYA surveillance data repository and the original fully identifiable encounter-level data received from the data sources will be eventually destroyed (including the clear-text PHI data needed for record linkage in accordance with the agreements put in place between SEARCH-DiCAYA and the data sources.

The University of Colorado will act as responsible stewards of patient data by maintaining, and whenever possible, strengthening the privacy and confidentiality of patient data stored in the SEARCH-DiCAYA surveillance system. Data stored in the SEARCH-DiCAYA surveillance system adhere with the HIPAA regulations specifying data files must be destroyed 7 years after IRB acknowledgement of study closure. De-duplicated, limited, encounter- and summary-level data will be generated on a periodic basis and submitted to the DiCAYA Coordinating Center at New York University (NYU) via the secure network as prescribed. Aggregate tables will be shared with external researchers in agreement with the CDC and other recipients. HIPAA compliant, fully de-identified data sets will be shared with external researchers upon approval of the DiCAYA governance committee.

NYU Data Management Plan (DMP)

I. Descriptions of data, access, storage, and sharing

Infrastructure Security

NYULH has privacy provisions that are strictly adhered to, including mandatory Security Awareness and HIPAA training of all employees (from custodial staff to Administration) and mandatory Good Clinical Practice and Protection of Human Subjects training for all staff involved in clinical research. All staff must sign a confidentiality agreement upon employment. NYULH meets or exceeds all HIPAA requirements. Faculty and staff are prohibited from keeping any clinical data on their desktop computers, including clinical research data; instead, it is stored on servers residing in the secured off-site data center. Transaction logs and database backup procedures allow the recreation of a clinical study at any point in time.

NYULH has strict security policies to ensure the privacy and confidentiality of data and guard against physical, accidental, or malicious loss of data or the hardware on which it resides. All resources, including web, database, and file servers are protected from outside intrusion by a firewall that blocks unauthorized access to the LAN by any unauthorized user originating from the Internet, using a sophisticated combination of secure application proxies and packet filtering. The NYU-managed network drive resides in our secured off-site data center. NYU Langone's high-performance off-site data center uses a network design based on Cisco Systems' Nexus data center switches. Data security is integrated into the network design by segmenting web tier, application tier, and database tier servers. A next-generation firewall system from Palo Alto Networks enables access control based on applications and user profiles; this granular, flexible security policy engine

simultaneously enables better services and better security. NYU's data center sub-host provider is Sungard. We provide a copy of our Service Organization Control (SOC) report attesting that we are operating under Statement on Standards for Attestation Engagements (SSAE) 18. An SSAE 18 attestation is similar to a FISMA attestation, except it is not assessed from a federal perspective.

Internal network security is maintained through Active Directory authentication. Intrusion detection software is employed to scan for attempted break-ins. User IDs and passwords are assigned and controlled as per SOP, and users are required by the system to change their passwords regularly. Access to clinical trials or other sensitive data is strictly limited and is granted only by the Director of the Technical Support Unit. Such access is controlled by easily identifiable group policies. Password policies prevent the use of repeat or similar passwords, and strong passwords are required and enforced. Three consecutive password failures cause the account to be locked until cleared by an NYULH IT Security Administrator. A single User ID and password is used for all server access, including file servers and database access. Users are instructed not to disclose their passwords and shared accounts where two or more individuals use the same login are prohibited. All databases that contain patient or clinical trial participant information are audited for unexpected database access and data changes.

Data Model & File Transfer

A modified version of the PCORnet Common Data Model version 6.0 will be used for submission of the limited dataset on cases meeting the computable phenotype definitions and to store and integrate these data in the central NYU-housed DiCAYA Network Database. A separate database will be created to store the limited datasets on the sample of wide-net cases that are selected for chart review validation.

Participating sites will transfer their research datasets as .txt files to internal NYU Langone Health data storage via GlobalScape Enhanced File Transfer (EFT), a secure FTP platform for data transfer. EFT Enterprise secures, manages, and tracks data transferred between people and applications, both inside and outside an organization. EFT can reduce complexity of the file transfer infrastructure, increase operational efficiency, and protect important data.

NYU will provide sites with their own login credentials to sftp://eftpub.nyumc.org, the NYU Langone GlobalScape site. DiCAYA sites can opt to use ssh-key for authentication instead of a password. Sites will be required to encrypt their files with Open PGP encryption for all data transfers involving PHI.

Once files arrive at the EFT server, rules are set i to immediately transfer those files to a secure internal NYU Langone server where access is limited to a select group of NYU users tasked with loading the data into the Hadoop database. Upon transfer, files are deleted from the EFT server.

Data Storage

All research data will be maintained or delivered as structured data in machine-readable data files (e.g., CSV, JSON, and XML) and will be stored on an NYU-managed network drive. DiCAYA data will be stored on the Hadoop Big Data platform hosted in the NYU Langone data center, using disk space allocated to Dr. Divers and Dr. Thorpe specifically for this project.

All data in the Hadoop platform are *encrypted at rest* and in *transition* using AES-CTR 256-bit encryption, TLS 1.2. Users are authenticated against NYU Langone identity services (Microsoft Active Directory) and can only access data if they are already on the NYU Langone network and use their network credentials. The Hadoop platform is integrated with Active Directory, and data authorization is based on the user's Kerberos ID and Active Directory group. For authorization, we use Role-Based Access Control with Apache Sentry. This provides fine-grained access to data accessible using schemas. Schemas are data structures described by the Apache Hive Metastore.

Data Access

The Hadoop data lake can only be accessed within the NYULH network and by authorized NYULH personnel. Within NYULH, access to the databases and network drives containing the research data for DiCAYA is controlled and managed by Dr. Divers and Dr. Thorpe. Access is granted only to those users who have a functional role in the study or the study database system. Dr. Divers and Dr. Thorpe have identified who will have access to the database, type of access permissions, and any specific user restrictions as appropriate. Upon notification that a user is no longer working on the study, the PI or designee will inform IT security to remove the user's access. If a user is inactive for longer than six months per system security requirements, the user will be inactivated.

Data Sharing

Summary statistics (i.e., counts, rates) by single year of age, race/ethnicity, sex, and geographic location will be shared with the DiCAYA Network sites and CDC via secure file transfer protocol (SFTP) under this FOA. Counts under ten will be shown as <10 instead of the exact number to protect privacy further. In addition, DiCAYA Network sites will be provided with their own site data with all computable phenotypes and derived variables generated at the CoC. The NYU CoC is also prepared to support targeted analyses from external stakeholders if deemed a priority by the Steering Committee. With the Steering Committee's approval, summary statistics (i.e., counts, rates) from these analyses may be shared with the stakeholders. NYU agrees to take care that all uses of DiCAYA Data, including but not limited to the creation of subsets, disclosure to authorized users, and publications, will conform to all requirements of 45 CFR § 164.514 (Privacy Rule).

II. Data Standards

Codebooks for harmonized datasets will be created and stored alongside data files. Codebooks will be shared with the sites through a common shared file directory (Google Drive), accessible through the study website. The NYU CoC will post the study protocol, manual of operations, training manuals, and related documents on the study website. The NYU CoC will share de-identified datasets and encourage investigators from outside the study to write papers and propose new ancillary studies.

NYU CoC DataCore investigators are responsible for managing, cleaning, removing duplicates, and correcting classification errors in all datasets. Data will be analyzed and merged using standard SQL and other statistical and analytics tools, including SAS, *R*, and GIS. Resultant datasets in

different formats will be linked by file naming conventions. Substantive metadata and source code will be generated during this process and will be stored with the data.

All data management activities of the project will conform to best practices and standards of NYU Langone: http://www.med.nyu.edu/irb/.

III. Archival and Long-Term Preservation of Data

All completed analyses, derived datasets, and metadata will be digitally archived through NYULH's network in formats that conform to the data storage standards of NYU Langone Medical Center Information Technology (MCIT) and the institutional Policy on Retention of and Access to Research Data.

VI. RECRUITMENT AND RETENTION PLAN

Protection of Human Subjects

Human Subjects Involvement, Characteristics, and Design

This integrated surveillance approach to observational, secondary data collection will involve ascertainment of individuals < 46 years of age with newly diagnosed or previously diagnosed with diabetes annually for the period 2020-2025. No direct patient contact will be involved, proposed data surveillance system will be constructed through the cross-linkage of multiple pre-existing electronic data sources which contain information on diabetes diagnosis, health care encounters, morbidity, and mortality related to diabetes and the integration of algorithms with an overall approach that includes of semi-automated incremental record linkage of diabetes-related healthcare encounters on each case across different healthcare systems and targeted chart review to ensure accurate estimates. Potentially eligible subjects will be identified based on the presence of a diabetes qualifying event during the surveillance periods which includes diabetes-related laboratory testing results, prescription of anti-hyperglycemic medication, or diabetes diagnoses. Actual eligibility as a diabetes case will be determined from extensive relevant EHR data dating back to 2012 received on potential cases with a diabetes qualifying event suspected of having diagnosed diabetes. In addition to the inclusion criteria of a diabetes-qualifying event, all cases will be age eligible (0-45 years) and reside in the state of Colorado at the time of the encounter. Exclusion criteria include individuals with gestational diabetes mellitus only and cases determined to be prisoners, wards of the state, or other institutionalized individuals. For case ascertainment, the SEARCH-DiCAYA case source network will include 8 hospitals: SCL Health, Children's Hospital of Colorado (CHCO), Barbara Davis Center (BDC), University of Colorado Health (UCH), Denver Health and Hospital Authority, Centura Hospitals, Boulder Community Hospitals and Valley-Wide Health Systems; 3 outpatient practices and clinics: Barbara Davis Center, Pediatric Endocrine Associates, and Rocky Mountain Pediatric Endocrinology; and one large data warehouses: Health Data Compass, which provides cases from UCH, CHCO, and the BDC.

Based on previous participation in the SEARCH study, we anticipate approximately 2089 incident and prevalent cases will be enrolled between 2020-2025, but actual enrollments may differ from this projection.

Since the focus of this study is to learn more about the impact of diabetes on people who are less than 46 years of age at the time of diagnosis, this study includes infants, children, adolescents, and young adults. This study does not involve fetuses, neonates, prisoners, or institutionalized individuals. Females who are determined to only have diabetes during a pregnancy will not be included.

Study Procedures, Materials, and Potential Risks

The study procedures and materials outlined below explain the sources and components included in structured datasets received from partnering institutions, the need for and procedures surrounding limited medical record review, and identifying the potential risks associated with this project.

Structured Dataset from Electronic Medical Record The sources of primary case finding for the SEARCH-DiCAYA surveillance system will expand on the well-established network of pediatric endocrinology clinics, hospitals and outpatient networks in Colorado established during the SEARCH study, presented in Table 8.

Table 9: Colorado Networks for case ascertainment
Colorado Sources of Cases
Hospitals, Locations (N in system)
SCL Health (formerly Exempla Hospitals) (N=4) statewide 💻
Denver Health (1 hospital and 8 primary care clinics with 1 integrated EHR), statewide ♥ 💻
Centura Hospitals (N=10) statewide 💻
Boulder Community Hospital 💻
BDC Adult Clinic (Aurora with outreach clinics) ♥ 💻
Valley Wide Health Systems ♥ 💻
CO out-patient sources
BDC (Aurora with outreach clinics) ♥ 💻
PEA (Denver with outreach clinics; for patients 0-25)
RMPE Endocrinologists (0-22) and ST. Mary's GJ of the Western Slope ♥
Data Warehouse Systems
Health Data Compass 💻 (includes adults from UCH, CHCO, and BDC)
Colorado All Payer Claims Database 💻
Secondary Enrichment Sources
Colorado Department of Public Health and Environment (CDPHE)

♥ *Minority, low income, medically underserved populations;* 💻 *EHR for case review available*

Each data partner will be asked to provide a structured dataset containing all relevant encounters going back to 2012 for each algorithm-identified probable case with a diabetes qualifying event during the 5-year period preceding the current surveillance year (for example in surveillance year 2020, the dataset would include 2015-2020; for surveillance year 2017, the dataset would include 2012-2017). The variables for the relevant healthcare encounters will be grouped into the following domains: diagnostic codes, encounter dates, laboratory measurements, clinical characteristics, demographics, medications, and personal identifying information (for de-duplication purposes). If the data partner is uncomfortable exchanging personal identifying information, analytic support will be provided to implement a hashing procedure to protect individual privacy while allowing the SEARCH-DiCAYA team to link patient encounter data with both the existing SEARCH Diabetes Registry and across data sources. Incremental record linkage will be used to reconcile, identify, and link potential duplicate cases identified across data sources and within the existing SEARCH Registry and across data sources to produce accurate estimates of incident and prevalent cases of diabetes. It is anticipated that a significant proportion the youth identified by the SEARCH-DiCAYA algorithm will have previously been identified by SEARCH and contained in the Registry if diagnosed prior to 2020. The incremental record linkage strategy will be implemented using a hybrid deterministic and probabilistic approach, designed to be both efficient and increase the accuracy of linkage for cases with missing data in key personal health identifiers. Newly identified cases (not determined to have a match in the existing SEARCH Registry) will be assigned a unique patient identifier and evaluated for additional diabetes-related encounters across data sources.

Medical Record Review Limited medical record information will be collected for subset of cases that fall into categories of the algorithm that have lower sensitivity and positive predicted value (PPV). To improve the overall performance of the ICD-10 based algorithms to determine diabetes type study staff will perform a targeted chart review of each newly identified probable diabetes cases identified by the ICD-10 based algorithm as undetermined and on a subset of other cases for validation purposes to confirm the presence or absence of diabetes, confirm diabetes type, and or to confirm date of diagnosis. Study staff will abstract information from the medical record for the purposes of ascertaining diabetes type and date on onset from the period of the first diabetes ICD-10 code to 6 months after this date. Information reviewed in the chart will include: 1) clinical notes, 2) results of diabetes autoantibody measurement (GAD65/GAA, IA2/ICA512, ICA, IAA, and ZnT8), 3) height and 10) weight (closest to diagnosis), 4) whether the participant ever used insulin, 6) whether insulin was discontinued, 7) presence of acanthosis nigricans, and 8) whether DKA was noted (with dates, bicarbonate, pH, and glucose values).

A structured data collection form will be completed by the data abstractor with fields for determined diabetes type and onset date.

Data Transmission, Security and Storage of PHI

Datasets from each of our partners will be transmitted to the LEAD Center using SSL (Secure Sockets Layer) certification to ensure a secure file transfer on an encrypted connection, through

encrypted email, or by mail on encrypted CDs. The LEAD Center will house the surveillance data repository of fully-identifiable patient level data on a comprehensive password-protected HIPAA-compliant database on a secure internal server. Access will be restricted to the Principal Investigator and core study personnel only. Upon receipt of each data set, study staff will give a randomly assigned unique ID number which refers to both the study record and the data source. All personal health identifiers (PHI) will then be stripped away from the primary dataset for storage and the PHI-ID link files will be maintained in a separate password-protected folder on the secure internal server. The PHI-ID link files will only be accessed for the purpose of record linkage as necessary.

Description of Risks and Justification of Study Approach

This is a no-contact analysis of secondary administrative data and historical electronic medical data. No treatment, intervention or procedures will be administered by the study. The purpose of the study is public health surveillance only. Confidentiality loss is the sole risk of this study. A breach of database security could result in a loss of confidentiality that could potentially result in psychological stress. It is unlikely that economic harm would ensue given the nature of the information. However, this risk is minimized in several ways described below.

Adequacy of Protection Against Risk

Informed Consent and Assent

Informed consent and assent will not be obtained for this no-contact analysis of secondary administrative data and historical electronic medical data. We will seek both a HIPAA Waiver and a Waiver of Consent to receive the data from our research partners. It would be impracticable to obtain consent from the entirety of the cases with diabetes and omitting relevant cases would bias results. We expect to submit the protocol, if awarded, as expedited for first review and exempt after the appropriate waivers are granted. We would put appropriate Business Associates Agreements and Data Use Agreements in place for our staff to receive full PHI from our collaborators and transmit limited data sets with limited data going to the coordinating center funded in Component C, New York University (NYU).

Protections against Risk

All unique identifiers will be used solely in the record-linkage phase where we identify individuals who are in one or more of the data sources. Once linkage has been accomplished, new study ID numbers will be assigned, and the PHI variables will be removed. The file linking PHI to the study ID will be maintained for the duration of the surveillance system in a separate, password-protected, encrypted file at the LEAD Center on a secure server. Study personnel will not alter electronic health records in any way and will have read-only access to hospital systems when limited medical record review is required. No findings or study data will ever be put into an individual's medical record, and data will not be available to employers, individuals, or other outside parties except as mandated by law, or for research purposes upon completion of all IRB, ethics, and review procedures. A certificate of confidentiality will be obtained as a further method of protection for study participants. Given the specific data requested from collaborators and the absence of direct participant contact in any way, we do not expect to have incidental findings. Nothing about the proposed research would make such a discovery possible.

Vulnerable subjects

Since the focus of part of this study is to learn more about the impact of diabetes on people who are less than 18 years of age at the time of diagnosis, this study includes infants, children, adolescents, and teens. This study does not involve fetuses, neonates, prisoners, or institutionalized individuals. Females who are determined to only have diabetes during a pregnancy will not be included. The protections outlined above are designed to protect the confidentiality of the minor children identified for this work. All other vulnerable populations are expressly excluded from the proposed research.

Potential Benefits of the Proposed Research to Human Subjects and Others

There are no direct benefits to study participants. Participation in this study may result in potential benefits to society. This is a large, state-wide component of a larger national study that will be well-represented by young people from a variety of racial/ethnic backgrounds. The information obtained in this study will help clinicians to better understand the characteristics of various types of diabetes, the frequency of co-morbidities and complications associated with diabetes, and the clinical impact diabetes has on the lives of these young people. This information will also be important in the planning of the distribution of medical and financial resources for the care of young people with diabetes in the future. Potential risks to study participants are minimal and reasonable in relation to the importance of the knowledge that is expected to be gained from this study since the only possible risk is a breach of confidentiality, which we have extensive and detailed methods to protect against, the possible benefits to society outweigh the possible risks.

Importance of the Knowledge to be Gained

Diabetes is the third most common chronic disease of childhood and adolescence. In the past, childhood diabetes was thought to consist almost exclusively of Type 1 diabetes. Over the past two decades, however, an increasing number of cases of Type 2 diabetes have been reported in children. Overall, the total number of diabetes cases affecting people less than 46 years of age is increasing over time.

This is a large, state-wide component of a larger national study that includes youth from diverse racial/ethnic and socioeconomic backgrounds. The information obtained in this study will help clinicians to better understand the characteristics of various types of diabetes, and the identification of complications of diabetes or an increased risk for developing complications. This information will also be important in the planning of the distribution of medical and financial resources for the care of young people with diabetes in the future.

Potential risks to study participants are minimal and reasonable in relation to the importance of the knowledge that is expected to be gained from this study since the only possible risk is a breach of confidentiality, which we have extensive and detailed methods to protect against, the possible benefits to society outweigh the possible risks.

VII. LOCAL DATA AND SAFETY MONITORING PLAN

Maintaining the accuracy and security of our data is vitally important. The data that will be produced as part of this project include:

- Local SEARCH-DiCAYA Diabetes Surveillance Database: an encounter and summary-level database of individuals with diabetes aged 0-17 and 18-45 in Colorado during each year of case ascertainment. The minimum amount of PHI necessary to uniquely identify cases and record necessary demographics will be locally maintained in a secure, password protected database.
- Clear-text PHI needed for record linkage and de-duplication of diabetes cases within and across data sources
- Limited SEARCH-DiCAYA Diabetes Surveillance Database: The limited SEARCH-DiCAYA Diabetes Surveillance Database will include limited information necessary to be made available to the NYU coordinating center, the study sponsors, and/or researchers in need of access to de-identified data. Both the local SEARCH-DiCAYA Diabetes Surveillance Database, with identifiers, and the limited SEARCH-DiCAYA Diabetes Surveillance Database would be password protected with user-based role accessibility and housed on the University of Colorado's secure Isilon server that is HIPAA compliant. The storage unit is approved by the security and compliance team for use, meets certain HIPAA and FERPA requirements that govern data integrity and is safeguarded by multi-layer firewall, intrusion, ransomware and anti-virus protections.

Data will be collected according to strict data standards to ensure reliability and reproducibility following both local and national SEARCH-DiCAYA protocols. In addition, a complete description of the methods of data collection, data dictionaries, and potential limitations of the data will be documented. The local SEARCH-DiCAYA Diabetes Surveillance Database will be developed by 12/31/2021. The clear-text PHI data will be received from primary case-finding data sources annually by May 31 starting in 2021.

The University of Colorado will house the local SEARCH-DiCAYA Diabetes Surveillance Database and the limited diabetes surveillance data repository in a password-protected Microsoft Access database on a secure internal server at the University. The University of Colorado computers able to access the local SEARCH-DiCAYA Diabetes Surveillance Database will be stored in locked offices. Key access to the offices will be restricted to the research team and the offices will always be locked when not occupied by the project personnel. The computers of the University of Colorado research team are on a segregated network utilized specifically for data that falls under the security rules of HIPAA. The University of Colorado uses network segregation as means of data separation as per HIPAA requirements. This network is disconnected from the standard, public University network using firewalling and routers. Additionally, the University of Colorado uses a system of "access control" for certain folders that are located within the HIPAA network that house the local SEARCH-DiCAYA Diabetes Surveillance Database and the SEARCH-DiCAYA data repository on the secure internal server. The folders are access restricted. Permissions are organized and granted by the IT department. Access to individual files and folders are assigned unique permissions stored in the Active Directory. The University of Colorado's IT team will authorize members of the research team to have access to a specific folder housing the project data.

Access to this folder is restricted to the research team, and users require passwords to access this folder. The passwords are a nonsensical combination of numbers and letters, changed on a regular schedule, never repeated, and stored away from the computer.

For the clear-text PHI data, additional security measures will be implemented. Briefly, we will create a virtual machine that is not an internet-facing server using a full-blown jump box desktop. For all technical and security standpoints, it will meet every detail specified in the Security Plan. The Privacy Rule permits assigning to, and retaining with, the health information a code or other means of record identification if that code is not derived from or related to information about the individual and could not be translated to identify the individual. The Security Rule operationalizes the protections afforded in the Privacy Rule by establishing standards for addressing the technical and non-technical safeguards that organizations must have in place to protect the privacy of individuals' PHI. The project team will implement and oversee policies, procedures, and technologies that are appropriate for their identified risks to PHI. The University of Colorado will remove the all PHI, or any other means of record identification, prior to uploading the data into the SEARCH-DiCAYA surveillance system so that there will be no method to re-identify the data. Only the limited data will be transferred to the SEARCH-DiCAYA surveillance data repository and the original fully identifiable encounter-level data received from the data sources will be eventually destroyed (including the clear-text PHI data needed for record linkage) in accordance with the agreements put in place between SEARCH-DiCAYA and the data sources.

The University of Colorado will act as responsible stewards of patient data by maintaining, and whenever possible, strengthening the privacy and confidentiality of patient data stored in the SEARCH-DiCAYA surveillance system. Data stored in the SEARCH-DiCAYA surveillance system adhere with the HIPAA regulations specifying data files must be destroyed 7 years after IRB acknowledgement of study closure.

De-duplicated, approved limited, encounter- and summary-level data will be generated on a periodic basis and submitted to the NYU CoC and/or the CDC as required via the secure network as prescribed. Aggregate tables will be shared with external researchers in agreement with the CDC and other recipients. Limited data sets will not be shared with external researchers.

VIII. REFERENCES

1. Pettitt DJ, Talton J, Dabelea D, Divers J, Imperatore G, Lawrence JM, et al. Prevalence of diabetes in U.S. youth in 2009: the SEARCH for diabetes in youth study. Diabetes Care 2014;37(2):402-8.
2. Dabelea D, Mayer-Davis EJ, Saydah S, Imperatore G, Linder B, Divers J, et al. Prevalence of type 1 and type 2 diabetes among children and adolescents from 2001 to 2009. JAMA 2014;311(17):1778-86.
3. Mayer-Davis EJ, Dabelea D, Lawrence JM. Incidence Trends of Type 1 and Type 2 Diabetes among Youths, 2002-2012. N Engl J Med 2017;377(3):301.

4. Patterson CC, Harjutsalo V, Rosenbauer J, Neu A, Cinek O, Skrivarhaug T, et al. Trends and cyclical variation in the incidence of childhood type 1 diabetes in 26 European centres in the 25 year period 1989-2013: a multicentre prospective registration study. Diabetologia 2019;62(3):408-417.
5. Group DP. Incidence and trends of childhood Type 1 diabetes worldwide 1990-1999. Diabet Med 2006;23(8):857-66.
6. Ye Q, Fu JF. Paediatric type 2 diabetes in China-Pandemic, progression, and potential solutions. Pediatr Diabetes 2018;19(1):27-35.
7. Wu H, Zhong J, Yu M, Wang H, Gong W, Pan J, et al. Incidence and time trends of type 2 diabetes mellitus in youth aged 5-19 years: a population-based registry in Zhejiang, China, 2007 to 2013. BMC Pediatr 2017;17(1):85.
8. Pinhas-Hamiel O, Zeitler P. Acute and chronic complications of type 2 diabetes mellitus in children and adolescents. Lancet 2007;369(9575):1823-1831.
9. Dart AB, Sellers EA, Martens PJ, Rigatto C, Brownell MD, Dean HJ. High burden of kidney disease in youth-onset type 2 diabetes. Diabetes Care 2012;35(6):1265-71.
10. Tryggestad JB, Willi SM. Complications and comorbidities of T2DM in adolescents: findings from the TODAY clinical trial. J Diabetes Complications 2015;29(2):307-12.
11. Dabelea D, Knowler WC, Pettitt DJ. Effect of diabetes in pregnancy on offspring: follow-up research in the Pima Indians. J Matern Fetal Med 2000;9(1):83-8.
12. Dabelea D, Mayer-Davis EJ, Lamichhane AP, D'Agostino RB, Jr., Liese AD, Vehik KS, et al. Association of intrauterine exposure to maternal diabetes and obesity with type 2 diabetes in youth: the SEARCH Case-Control Study. Diabetes Care 2008;31(7):1422-6.
13. Child and Adolescent Health Measurement Initiative. Prevalence of current diabetes, National Survey of Children's Health, 2017-18. Data Resource Center for Child and Adolescent Health supported by the U.S. Department of Health and Human Services, Health Resources and Services Administration (HRSA), Maternal and Child Health Bureau (MCHB). Retrieved [12/06/19] from [www.childhealthdata.org].
14. Petitti DB, Klingensmith GJ, Bell RA, Andrews JS, Dabelea D, Imperatore G, et al. Glycemic control in youth with diabetes: the SEARCH for diabetes in Youth Study. J Pediatr 2009;155(5):668-72 e1-3.
15. Skinner AC, Skelton JA. Prevalence and trends in obesity and severe obesity among children in the United States, 1999-2012. JAMA Pediatr 2014;168(6):561-6.
16. Imperatore G, Boyle JP, Thompson TJ, Case D, Dabelea D, Hamman RF, et al. Projections of type 1 and type 2 diabetes burden in the U.S. population aged <20 years through 2050: dynamic modeling of incidence, mortality, and population growth. Diabetes Care 2012;35(12):2515-20.
17. Rewers A, Klingensmith G, Davis C, Petitti DB, Pihoker C, Rodriguez B, et al. Presence of diabetic ketoacidosis at diagnosis of diabetes mellitus in youth: the Search for Diabetes in Youth Study. Pediatrics 2008;121(5):e1258-66.
18. Dabelea D, Rewers A, Stafford JM, Standiford DA, Lawrence JM, Saydah S, et al. Trends in the prevalence of ketoacidosis at diabetes diagnosis: the SEARCH for diabetes in youth study. Pediatrics 2014;133(4):e938-45.

19. Jensen ET, Stafford JM, D'Agostino RB, Jr., Dolan L, Lawrence JM, Marcovina S, et al. Rapid increase in prevalence of DKA at diagnosis among youth with type 1 diabetes: the SEARCH for Diabetes in Youth study In: ESAD; 2019; Barcelona, Spain; 2019.
20. Desai D, Mehta D, Mathias P, Menon G, Schubart UK. Health Care Utilization and Burden of Diabetic Ketoacidosis in the U.S. Over the Past Decade: A Nationwide Analysis. Diabetes Care 2018;41(8):1631-1638.
21. Duca LM, Reboussin BA, Pihoker C, Imperatore G, Saydah S, Mayer-Davis E, et al. Diabetic ketoacidosis at diagnosis of type 1 diabetes and glycemic control over time: The SEARCH for diabetes in youth study. Pediatr Diabetes 2019;20(2):172-179.
22. Ziegler AG, Nepom GT. Prediction and pathogenesis in type 1 diabetes. Immunity 2010;32(4):468-78.
23. Group TS. Effects of metformin, metformin plus rosiglitazone, and metformin plus lifestyle on insulin sensitivity and beta-cell function in TODAY. Diabetes Care 2013;36(6):1749-57.
24. Group TS, Zeitler P, Hirst K, Pyle L, Linder B, Copeland K, et al. A clinical trial to maintain glycemic control in youth with type 2 diabetes. N Engl J Med 2012;366(24):2247-56.
25. American Diabetes A. Diagnosis and classification of diabetes mellitus. Diabetes Care 2010;33 Suppl 1:S62-9.
26. Imperatore G, Mayer-Davis EJ, Orchard TJ, Zhong VW. Prevalence and incidence of type 1 diabetes among children and adults in the United States and comparison with non-U.S. countries. In: Cowie CC, Casagrande SS, Menke A, Cissell MA, Eberhardt MS, Meigs JB, et al., editors. Diabetes in America. Bethesda, MD.: National Institutes of Health; 2018. p. 2-1 - 2-7.
27. Wilkin TJ. The accelerator hypothesis: a review of the evidence for insulin resistance as the basis for type I as well as type II diabetes. Int J Obes (Lond) 2009;33(7):716-26.
28. Group SS. SEARCH for Diabetes in Youth: a multicenter study of the prevalence, incidence and classification of diabetes mellitus in youth. Control Clin Trials 2004;25(5):458-71.
29. Hamman RF, Bell RA, Dabelea D, D'Agostino RB, Jr., Dolan L, Imperatore G, et al. The SEARCH for Diabetes in Youth study: rationale, findings, and future directions. Diabetes Care 2014;37(12):3336-44.
30. Crume TL, Hamman RF, Isom S, Talton J, Divers J, Mayer-Davis EJ, et al. Factors influencing time to case registration for youth with type 1 and type 2 diabetes: SEARCH for Diabetes in Youth Study. Ann Epidemiol 2016;26(9):631-7.
31. Lawrence JM, Black MH, Zhang JL, Slezak JM, Takhar HS, Koebnick C, et al. Validation of pediatric diabetes case identification approaches for diagnosed cases by using information in the electronic health records of a large integrated managed health care organization. Am J Epidemiol 2014;179(1):27- 38.
32. Zhong VW, Pfaff ER, Beavers DP, Thomas J, Jaacks LM, Bowlby DA, et al. Use of administrative and electronic health record data for development of automated algorithms for childhood diabetes case ascertainment and type classification: the SEARCH for Diabetes in Youth Study. Pediatr Diabetes 2014;15(8):573-84.

33. Zhong VW, Obeid JS, Craig JB, Pfaff ER, Thomas J, Jaacks LM, et al. An efficient approach for surveillance of childhood diabetes by type derived from electronic health record data: the SEARCH for Diabetes in Youth Study. J Am Med Inform Assoc 2016;23(6):1060-1067.
34. Vanderloo SE, Johnson JA, Reimer K, McCrea P, Nuernberger K, Krueger H, et al. Validation of classification algorithms for childhood diabetes identified from administrative data. Pediatr Diabetes 2012;13(3):229-34.
35. Paul MM, Greene CM, Newton-Dame R, Thorpe LE, Perlman SE, McVeigh KH, et al. The state of population health surveillance using electronic health records: a narrative review. Popul Health Manag 2015;18(3):209-16.
36. Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PR, Bernstam EV, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. Med Care 2013;51(8 Suppl 3):S30-7.
37. Terry AL, Chevendra V, Thind A, Stewart M, Marshall JN, Cejic S. Using your electronic medical record for research: a primer for avoiding pitfalls. Fam Pract 2010;27(1):121-6.
38. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc 2013;20(1):144-51.
39. Brown JS, Kahn M, Toh S. Data quality assessment for comparative effectiveness research in distributed data networks. Med Care 2013;51(8 Suppl 3):S22-9.
40. Ahmad FS, Chan C, Rosenman MB, Post WS, Fort DG, Greenland P, et al. Validity of Cardiovascular Data From Electronic Sources: The Multi-Ethnic Study of Atherosclerosis and HealthLNK. Circulation 2017;136(13):1207-1216.
41. Klompas M, Cocoros NM, Menchaca JT, Erani D, Hafer E, Herrick B, et al. State and Local Chronic Disease Surveillance Using Electronic Health Record Systems. Am J Public Health 2017;107(9):1406- 1412.
42. Casey JA, Schwartz BS, Stewart WF, Adler NE. Using Electronic Health Records for Population Health Research: A Review of Methods and Applications. Annu Rev Public Health 2016;37:61-81.
43. Ong T, Duca LM, Kahn MG, Crume TL. A hybrid approach to record linkage using a combination of deterministic and probabilistic methodology. Journal of the American Medical Informatics Association 2020, Accepted, In Press.
44. Wells B, Lenior K, Wagenknecht L, Mayer-Davis E, Lawrence JM, Dabelea D, et al. Detection of Diabetes Status and Type in Youth Using EMRs. In: 79th Annual American Diabetes Association; 2019 June 2019; San Francisco, CA; 2019.
45. Duca LM, Pyle L, Khanna A, Ong T, Kahn MG, DiGuiseppi C, et al. Estimating the Prevalence of Congenital Heart Disease among Adolescents and Adults in Colorado Adjusted for Incomplete Case Ascertainment. American Heart Journal 2020;Accepted, In press.
46. Pasquali SK, Jacobs JP, Shook GJ, O'Brien SM, Hall M, Jacobs ML, et al. Linking clinical registry data with administrative data using indirect identifiers: implementation and validation in the congenital heart surgery population. Am Heart J 2010;160(6):1099-104.
47. Cox S, Martin R, Somaia P, Smith K. The development of a data-matching algorithm

to define the 'case patient'. Aust Health Rev 2013;37(1):54-9.
48. Jaro MA. Probabilistic linkage of large public health data files. Stat Med 1995;14(5-7):491-8.
49. Fellegi IS, A. A theory for record linkage. Journal of the American Statistical Association 1969;64(328):1183-1210.
50. Bruno G, LaPorte RE, Merletti F, Biggeri A, McCarty D, Pagano G. National diabetes programs. Application of capture-recapture to count diabetes? Diabetes Care 1994;17(6):548-56.
51. Hook EB, Regal RR. Capture-recapture methods in epidemiology: methods and limitations. Epidemiol Rev 1995;17(2):243-64.
52. Hamman RF, Dabelea D, Liese AD, et al. Estimation of Completeness of Case Ascertainment Using Capture-Recapture (Technical Report) [internet], 2013. Available from https://www.searchfordiabetes.org/public/dspPubs.cfm. Accessed December 2, 2019
53. American FactFinder, United States Census Bureau. State population characteristics 2010-2018. Available at: https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-detail.html. Accessed 12/2/2019
54. Bird SM, King R. Multiple Systems Estimation (or Capture-Recapture Estimation) to Inform Public Policy. Annu Rev Stat Appl 2018;5:95-118.
55. Gart JJ, Nam JM. Approximate interval estimation of the difference in binomial parameters: correction for skewness and extension to multiple tables. Biometrics 1990;46(3):637-43.
56. Nam JM. A simple approximation for calculating sample sizes for detecting linear trend in proportions. Biometrics 1987;43(3):701-5.
57. Rodriguez BL, Dabelea D, Liese AD, Fujimoto W, Waitzfelder B, Liu L, et al. Prevalence and correlates of elevated blood pressure in youth with diabetes mellitus: the SEARCH for diabetes in youth study. J Pediatr 2010;157(2):245-251 e1.
58. Kershnar AK, Daniels SR, Imperatore G, Palla SL, Petitti DB, Pettitt DJ, et al. Lipid abnormalities are prevalent in youth with type 1 and type 2 diabetes: the SEARCH for Diabetes in Youth Study. J Pediatr 2006;149(3):314-9.
59. Maahs DM, Snively BM, Bell RA, Dolan L, Hirsch I, Imperatore G, et al. Higher prevalence of elevated albumin excretion in youth with type 2 than type 1 diabetes: the SEARCH for Diabetes in Youth study. Diabetes Care 2007;30(10):2593-8.
60. Mayer-Davis EJ, Davis C, Saadine J, D'Agostino RB, Jr., Dabelea D, Dolan L, et al. Diabetic retinopathy in the SEARCH for Diabetes in Youth Cohort: a pilot study. Diabet Med 2012;29(9):1148- 52.
61. Jaiswal M, Lauer A, Martin CL, Bell RA, Divers J, Dabelea D, et al. Peripheral neuropathy in adolescents and young adults with type 1 and type 2 diabetes from the SEARCH for Diabetes in Youth follow-up cohort: a pilot study. Diabetes Care 2013;36(12):3903-8.
62. Urbina EM, Wadwa RP, Davis C, Snively BM, Dolan LM, Daniels SR, et al. Prevalence of increased arterial stiffness in children with type 1 diabetes mellitus differs by measurement site and sex: the SEARCH for Diabetes in Youth Study. J Pediatr 2010;156(5):731-7, 737 e1.

IX. **DiCAYA COVID Supplement Protocol**

Background

Emerging studies suggest that there is a complex relationship between SARS-CoV-2 infection and diabetes. Early on, diabetes was found to be a risk factor for hospitalization and death among people who become infected with the SARS-CoV-2 virus.[1,2] Infection with SARS CoV-2 worsens diabetes outcomes and glycemia, because of β-cell damage, cytokine-induced insulin resistance, hypokalemia, drugs used to treat COVID-19, and disruptions in needed healthcare.[3,4] Most of the research on COVID-19 and diabetes has focused on adult populations, where background rates of diabetes and its comorbidities are higher. However, preliminary evidence also suggests that COVID-19 infection worsens diabetes and COVID-19 outcomes among children with diabetes[5–7] For example, a meta-analysis on the incidence of diabetic ketoacidosis (DKA) among children with Type I diabetes before and during the COVID-19 pandemic identified that DKA risk, especially the risk of severe DKA, has increased significantly during the pandemic.[8]

Viral infection has long been considered a possible precipitating factor for new onset Type 1 diabetes[9,10] and perhaps Type 2 diabetes.[11] Many people suspect that SARS-CoV-2 infection is such a factor. Case reports, small case series, observational studies (and systematic reviews) of varying quality, and speculative pieces about potential mechanisms all have appeared.[12–21] Observational studies of this topic can be difficult to interpret because of ascertainment and other biases, and confounding. And during the past three years, health and health care and many other aspects of life have been disrupted, by the virus but also by the way that institutions and people have reacted to the pandemic.

To date, the possibility of an association between SARS-CoV-2 infection and incident diabetes seems more likely among adults and those with more substantial COVID illness. The evidence base is less developed among children and those with mild or asymptomatic infection. A recent meta-analysis of 14 studies from 2020 (age ranges varied) concluded that "COVID-19 survivors may be at increased risk for new-onset diabetes."[21] By contrast, a large, observational, reasonably well-done study in Scotland concluded that "Type 1 diabetes incidence in children increased during the pandemic" but suggested that "SARS-CoV-2 infection itself was not the cause of this increase."[18] A study in-progress in PEDSnet (interim data presented by CHOP at the 10/27/2022 DiCAYA meeting) suggested that SARS-CoV-2 infection is not associated with an increased incidence of diabetes in children.

The '*Assessing the Burden of Diabetes by Type in Children, Adolescents, and Young Adults'* (DiCAYA) Network aims to make a substantial contribution to the literature of SARS CoV-2 infection and new diabetes diagnosis risk. We share this motivation with Rubino et al., who established the CoViDiab registry to study whether the association is true, and, if yes, for which diabetes type(s).[10] Electronic health records (EHRs) provide a needed, time-sensitive approach. We propose leveraging large-volume EHR data across 8 US-based centers funded as part of the DiCAYA Network. Using a harmonized protocol and data sharing platform supported by a coordinating center, half of DiCAYA centers conduct diabetes surveillance in youth (0-17 yrs), and half conduct diabetes surveillance in young adults (18-45 yrs).

Our Specific Aims are:

AIM 1. To assess and improve data quality of SARS CoV-2 measures among children and young adults with prior SARS CoV-2 infection and/or SARS CoV-2 vaccination.

AIM 2. To adapt current DiCAYA computable phenotypes (designed to detect incident Type I and Type II diabetes for annual population estimation) for a multi-center retrospective cohort study of post-COVID diabetes risk in children and young adults to characterize post-infection onset time distributions.

AIM 3. To estimate relative and absolute associations between SARS CoV-2 infection and new diabetes diagnosis risk occurring between June 1, 2020 and December 31, 2022, among children (0-17) and young

adults (18-44), overall and by age group and sex. Specifically, we will compare the relative and absolute risk of new diabetes diagnoses (based on DiCAYA's working computable phenotype definition) in patients seen at least once in 2018 or 2019, comparing (a) patients with a record of a SARS-CoV-2 test or illness prior to December 31, 2021, and (b) a contemporary comparison group of patients with no record of SARS-CoV-2 infection.

Hypothesis: Adjusting for other covariates, children and young adults with documented SARS-CoV-2 infection will have a higher risk and burden of diabetes than those without evidence of SARS-CoV-2 infection.

Roles

Protocol development. To guide decision-making for these analyses, the Data Analysis Working Group (DAWG) has assigned development of a draft protocol to a subcommittee. This subcommittee will present aspects of the protocol to DAWG in October-November, 2022. DAWG comprises members from each Center, the Coordinating Center, and the Sponsor. DAWG will then present the draft protocol to the DiCAYA Steering Committee for a vote, ideally before the end of December, 2022. Non-consensus on the protocol is defined as anything with less than 80% support among the relevant Centers (i.e., among Component A Centers when voting on analytic approach for children, among all Centers for DUAs/data sharing), and does not refer to elements of the protocol that are not feasible based on local institutional policy. The CoC will be working with individual sites to address the latter. The DUA also specifies different levels of data sharing to accommodate sites with institutional policies that put restrictions on the volume and data types that can be shared with the CoC. After the protocol has been approved, maintenance of the protocol will be maintained by NYU CoC with DAWG oversight, and the subcommittee will no longer be active.

Center cohort construction and local analyses. Each Center will designate a local analyst to guide local propensity score estimation and analyses. While many other investigators and staff from each Center will be involved in this project, having these individuals identified will facilitate more rapid progress between meetings. The CoC will develop sample codes and table shells where applicable to ensure that all sites work from a common workplan.

CoC support. The NYU CoC will prepare guidance materials for each step of the project, including maintenance of this supplement protocol. All guidance materials will undergo DAWG review and final versions be shared with Centers. Where possible, the CoC will build on the DiCAYA-specific common data model (CDM) developed for the parent study to identify proposed data elements and structures, but the NYU CoC will not be providing ETL scripts directly for cohort construction. The NYU CoC will provide table shells and sample code to guide the iterative propensity score estimation process, as well as material to guide person-time construction and regression analyses.

Study Design (Cohort Definitions and Timeline)

Inclusion criteria and comparison populations. Local Centers will create propensity score-weighted cohorts for the primary analysis that include (children or young adults) with documentation of a COVID-19 infection and individuals without documentation of a COVID-19 infection. Definitions for the "exposed" and "unexposed" populations for the primary analysis are provided in Table 1. All individuals are required to have at least one health care encounter in 2018 or 2019. Membership-based sites may consider altering the eligibility criteria to require membership (instead of a health care encounter) at any time in 2018 and 2019, as well as active membership on June 1, 2020. Membership-based sites will provide the CoC and the Network with a table comparing the size of the potential sample under each eligibility definition (utilization- versus membership-based) to inform final Network decision on optimal

approach. If population size differences are relatively small, retaining consistency across sites may be preferred; if relatively large, membership eligibility may reduce important potential biases. All participants are also required to have a minimum of one health care encounter (in addition to the index encounter) during the follow-up period to ensure contribution of some person-time. This additional health care encounter may be a telehealth visit. The index encounter is the date of the first positive SARS CoV-2 test or first COVID-19 diagnosis code among the exposed group, and is the date of a health care encounter between June 1, 2020 and December 31, 2021 for the unexposed group.

Sites that are not able to include all eligible individuals in the analysis (due to data transfer restrictions, for example) may select all SARS-CoV-2 exposed individuals, along with a random subset of unexposed individuals. We propose that sites select at least 3 unexposed individuals for every exposed individual to ensure adequate sample size for a cohort-based analysis.

Exclusion criteria: Individuals with a prior history of diabetes (any diagnosis code or medication prior to the start of follow-up) will be excluded. Unexposed children will be excluded if they have any documentation of MIS-C diagnosis codes prior to or during the follow-up period. Individuals will be excluded if they are missing covariate information (with the exception of race/ethnicity). Individuals with documentation of a positive SARS CoV-2 test or SARS CoV-2 diagnosis code prior to June 1, 2020 will be excluded.

Local sites can use the inclusion/exclusion criteria and definitions for exposed/unexposed individuals (**Table 1**) to construct the local cohorts.

Table 1. Exposed/unexposed definitions for primary analysis

Exposed definition*	**Unexposed definition**	**Follow-up period**
Documentation of at least one positive SARS CoV-2 molecular [PCR]/antigen test **or** at least two COVID-19 diagnosis codes (e.g., U07.1) between June 1, 2020 and December 31, 2021** (and within three months of one another), or either of the following positive serology tests: • Nucleocapsid Ab (+) • Spike Ab (+) prior to vaccine availability: o Dec 16, 2020 for 16+ o May 12, 2021 for 12-15 o Nov 2, 2021 for 5-11	No documentation of a positive SARS CoV-2 molecular/antigen test and no record of any COVID-19 diagnosis codes prior to December 31, 2021	Through December 31, 2022

Assumptions: *Start with broader exposure definition including 2+ COVID diagnoses, but if Aim 1 assessment suggests we are capturing incidental non-infections (e.g. rule-out testing), drop. **We are requiring two codes based on experiences of RECOVER and elsewhere, which suggests that many times diagnosis codes are used for rule out diagnoses, exposures etc.

Proposed secondary analyses to refine cohort definitions:

- Revision of COVID-19 exposure definition (related to Aim 1).
    - Justification: There are multiple algorithms in the literature used to identify SARS-CoV-2 infections. Different methods may trade off sensitivity and specificity, and their performance may also vary over time.
    - Description: Sites will explore alternative definitions for identifying COVID-19 exposed/unexposed individuals. An alternative approach may include treating all ARIs in

children in early stages of the pandemic as COVID-related. Analytic model results will be compared using different exposure definitions.

- Stratification by pandemic wave.
    - Justification: The sensitivity/specificity of definitions for positive/presumed positive COVID-19 cases may vary across stages of the pandemic. Associations between COVID-19 and incident diabetes may also vary by variant.
    - Description: For sites with adequate sample size, cohort definitions (and analyses) will be stratified by variant wave (alpha, delta).
        - Proposed dates (to be verified with PEDSnet regarding whether dates should vary by region):
            - Alpha = 9/15/20-6/5/21
            - Delta = 7/18/21-12/4/21

- Test-negative design.
    - Justification: Unequal uptake of COVID testing resulting in differential ascertainment of either COVID-19 status or diabetes may introduce potential bias into the analysis.
    - Description: We will conduct a secondary analysis using a "test-negative" design, including only individuals with a PCR-positive SARS CoV-2 test or a PCR-negative SARS CoV-2 test. The index date for the test-negative design is the date of the first PCR-positive SARS CoV-2 test (exposed) and the date of a PCR-negative SARS CoV-2 test (unexposed).
        - Exposed group definition: Documentation of at least one positive SARS CoV-2 molecular test between June 1, 2020 and December 31, 2021.
        - Unexposed group definition: Documentation of at least one negative SARS CoV-2 molecular test between June 1, 2020 and December 31, 2021, and no documentation of a positive SARS CoV-2 molecular/antigen test or record of any COVID-19 diagnosis codes prior to December 31, 2021.

Timeframe. As noted in Table 1, the study period is June 1, 2020 – December 31, 2022. We exclude March-May 2020 due to insufficient testing capacity nationally.

Study Design (Analytic Approach)

AIM 1: Refining Exposure Definitions and Harmonization (SARS CoV-2 Exposures). To assess and improve data quality of SARS CoV-2 measures, we will follow an adjudication process led by the Data Analysis Working Group that builds upon extensive work already completed at several PEDSnet/PCORnet-affiliated institutions within DiCAYA. Inconsistencies between sites will be documented, and Network-wide decisions will be made on harmonizing and managing missing data. Data quality assessment and harmonization will occur in the first 4-6 months of the supplement grant year.

- Distributions across sites will be compared for the following components of the SARS CoV-2 infection exposure definition, and unusual distributions will be investigated:
    - Positive SARS CoV-2 PCR tests
    - Positive SARS CoV-2 antigen test
    - At least two COVID-19 diagnosis codes (e.g., U07.1) between June 1, 2020 and December 31, 2021, within three months of one another
    - Positive serology tests:
        - Nucleocapsid Ab (+)

- Spike Ab (+) prior to vaccine availability

AIM 2: Adapting Primary Outcome. For this analysis, we will be adapting DiCAYA diabetes definitions. Our primary outcome for the study is a new diagnosis of diabetes. The working computable phenotype (CP) definition for diabetes defined in DiCAYA's Protocol will be modified to address the specific research questions outlined in this protocol. While we will characterize all new diabetes diagnoses that occur after the index SARS CoV-2 test or COVID-19 diagnosis (date of first diagnosis code), the time window for the primary outcome will be set at least 30 days after the index date (excluding cases of transient, resolved hyperglycemia). We will also require only one new diabetes diagnosis code, compared to ≥2 outpatient diagnosis codes in the parent study.

- Secondary Outcome. Diabetic ketoacidosis (DKA) defined as E08.1, E09.1, E10.1, E11.1, or E13.1, coded during the observation period.
- Potential Ancillary Study. As noted, transient hyperglycemia is outside of the scope of the current study. Examining associations between SARS CoV-2 infection and transient hyperglycemia may be of interest in an ancillary study.

AIM 3: Associations between SARS CoV-2 infection and new Diabetes Diagnoses. Local Centers will construct eligible cohorts, estimate propensity weights, locally estimate time-to-event models, and transfer summary data to the CoC for meta-analysis. Throughout this process, the CoC will support local site cohort development and ensure consistent measure development, analysis, and approaches to data quality control.

An overview of the analytic process is as follows:

1. Sites construct local cohorts based on common inclusion/exclusion criteria and definitions
2. CoC prepares table shell including individual- and area-level covariates by COVID exposure status (see Table 2)
3. Sites share zip codes; CoC prepares area-level variables (based on ACS data and modified RUCA codes from Diabetes LEAD Network) centrally
4. Sites populate table shells using individual-level and linked area-level data
5. CoC compiles/disseminates summary tables to network
6. Sites generate person-time following standardized protocol (described below) and share with CoC for DiCAYA-wide unweighted Kaplan-Meier curve generation
7. Sites run unweighted Cox proportional hazard models and share results with CoC; CoC compiles/disseminates summary tables to network
8. Sites estimate propensity weights and apply weighted person-time to Cox models (e.g., using the R package 'MatchIt'). Censoring weights (using inverse probability weighting) will be developed as needed, based on the likelihood of censoring in the dataset. For example, if censoring is high (>10%), we propose to create 4 groups (based on exposure and censoring status) and either create independent censoring weights (to be multiplied by the propensity weights generated to address confounding), or, fit a 4-category logit model to generate joint exposure and censoring weights.
9. Sites share summary data with CoC for meta-analysis
10. Process repeated for sensitivity and secondary analyses, stratified analyses, etc.

Table 2 includes a table shell for sites to populate for key covariates using unweighted cohort data. Examining unweighted data will reveal the extent to which exposed/unexposed individuals vary

according to key characteristics, prior to propensity score estimation. This table can be populated using the 'MatchIt' package in R.

Table 2. Table shell comparing unweighted standardized mean differences for key covariates across COVID-19-exposed/unexposed cohorts (primary analysis)

	Means exposed	Means unexposed	Std. Mean Diff.
BMI (mean)			
Male (%)			
Non-Hispanic White (%)			
Non-Hispanic Black (%)			
…			

Propensity score estimation. Local propensity score estimation will be conducted by Centers. We recommend using the R package 'MatchIt', though other statistical packages may be used (e.g., the SAS macro, 'PSM'). To facilitate a smooth process, the CoC is proposing to pilot the propensity score estimation process with one or two Component A and Component B sites. The CoC will then share steps and suggested code for propensity score estimation across the full network. Individual sites will agree to not move forward with the propensity score estimation process until pilot testing is complete. The proposed process includes the following steps:

1. CoC lays out analytic approach (guided by DA Working Group)
2. Sites estimate propensity scores weights using logistic regression
3. CoC and sites review convergence, balance and weights. Sites share tables of standardized effect sizes and differences, unweighted and weighted (assess for SD<0.2 or another cut-off).
4. Sites re-run propensity score estimation if necessary to improve balance.
5. Sites calculate stabilized weights. Sites trim any excess weights (e.g. 95th percentile).
6. CoC refines process with pilot sites and shares with rest of sites, including analytic code examples
7. Sites run pre-prepared propensity score weighting strategy and share same results as above with CoC to compile and review with DA Working Group
8. Refine as needed

Example code and vignettes for R 'MatchIt' can be found here: https://cran.r-project.org/web/packages/MatchIt/vignettes/MatchIt.html

The goal will be to retain as similar a process as possible across Centers, barring necessary modifications to achieve a successful balanced cohort. The PS estimation process may be repeated for stratified analyses to achieve balance across strata.

Covariates. Table 3 includes key confounders to be balanced in the propensity score weighting process, including demographics, healthcare utilization, and other clinical and area-level covariates. For the primary analysis, index month represents the month of first positive SARS CoV-2 test (or first SARS CoV-2 diagnosis code) for the exposed group and month of encounter for the unexposed group. For the test-negative design, the index month is the month of the SARS-CoV-2 diagnostic test. Measures to include in the propensity score approach are based on known and hypothesized common causes and observed subgroup differences. The Network will also collaboratively refine and add sub-comparison groups as needed. The R package 'comorbidity' may be used to systematically explore weighting scenarios for comorbidity indices across sites. A data dictionary for proposed covariates will be provided.

Table 3. Key covariates to be included in propensity score approach

	Proposed form	Data source
Potential individual-level variables		
Age at index month	Categorical Children: TBD (0-9, 10-17) Young Adults: TBD (18-29, 30-45)	EHR
Sex	Categorical (Male, Female)	EHR
Race/ethnicity	Categorical (NH White, NH Black, Hispanic, Asian, Multiracial, NA/AI, Other, Unknown)	EHR
BMI (most recent measurement prior to or on index date)	Continuous	EHR
Total visits (all types) in 2018 and 2019	Continuous	EHR
Index month	Categorical	EHR
Comorbidity profile	Component A: PCMA: no chronic condition, noncomplex chronic condition, or complex chronic condition (PEDSnet)[22] Component B: Elixhauser comorbidity index (Assess feasibility)	EHR
Smoking status	Current smoker/not current smoker/missing	EHR
Medication history	TBD	EHR
Area-level variables (ZCTA)		
Percent of families with annual income < poverty level	Continuous	ACS
Percent of individuals over age 25 with < high school education	Continuous	ACS
Urban/rural status	Categorical (High density urban, low density urban, suburban/small town, rural)	Modified Rural-Urban Commuting Area (RUCA) codes from LEAD Network[23]

Person-time calculation. Before analyzing associations in Aim 3, we will characterize the time to diabetes diagnosis and DKA, with time to diagnosis defined as the difference between the index date and first diagnosis of diabetes or DKA occurrence. While causal associations are not yet ascertained and estimation of actual incubation periods is complicated by when patients seek healthcare and receive appropriate tests, this characterization will provide initial descriptive information to inform this question and will also be used to refine exclusion criteria for Aim 3 analysis. Person-time will be calculated through December 31, 2022 or censoring, defined as the first diagnosis of diabetes or DKA occurrence, death, or disenrollment (membership-based sites). Sites will calculate person-time locally, and send person-time data on exposed and unexposed cases (full weighted sample or random sample) to the CoC for a DiCAYA-wide Kaplan-Meier curve. This data transfer is covered by the current DUA.

Time-to-event analysis. Once propensity score-weighted cohorts have been constructed, local sites will estimate weighted time-to-event models. Cox regression models and gamma-frailty models will be used to estimate adjusted HRs for diabetes risk while adjusting for unobserved heterogeneity. Survival models will also generate excess burden per 10,000 [children/adults] based on the difference in survival

probability between groups and transformed as event rate difference, similar to what was completed in a recent study on post-COVID diabetes risk among adults.[5] Summary data from these models will be shared with the CoC. Table shells are provided in Tables 4 and 5.

Once the time to diabetes incidence has been defined, generation of the data needed to fill Table 4 is fairly straightforward, with R package *survival*. The CoC will generate the initial code, which will be shared with the sites. To understand the effect of weighting, cumulative incidence plots will be generated for the weighted and unweighted analyses. Table 5 will be formalized as the preliminary data become available, but we can envision an initial model testing for the effect of the SARS-CoV-2 infection on time to diabetes with a single predictor, and models with additional covariates using the doubly robust framework.

Table 4. Table of person-years for Kaplan-Meier plots

Time in months	Number at risk	Number of events	Number censored	Cumulative incidence
0				
1				
.				
.				
.				
24				

Table 5. Parameters from local site analyses and DiCAYA estimate from meta-analysis

		GEI	KAI	IUP	UCO (B)	LUR (B)	DiCAYA estimate
Model 1	β_0						
Model 2	β_0						
	β_1						
	β_2						

Proposed secondary analyses to explore biases associated with meta-analytic approach:

- Transfer of line-level data for analytic cohort to CoC for pooled analysis.
    - Justification: While the primary approach is a meta-analysis, meta-analytic methods may be biased when sample sizes are small and outcomes are rare. Pooled analysis at the CoC using line-level data could potentially shed light on any bias associated with the meta-analytic approach.
    - Description: For sites with this capability (where line-level data sharing is specified in DUAs), line-level data from weighted cohorts will be transferred to the CoC for a pooled analysis. Results will be compared between the pooled line-level analysis and a meta-analysis using summary data from the same sites.

Stratified analyses. To examine diabetes risk in sub-groups, a select number of stratified analyses will be conducted, conditional on covariates other than the subgroup. This may require re-estimation of propensity scores within strata to achieve homogenous subgroups with similar propensity scores.

Priority stratified analyses include: age group and sex.

Secondary stratified analyses include: race/ethnicity, geographic region, hospitalization status (possible subgroup comparisons within hospitalized patients include SARS CoV-2 infection vs. other respiratory infections and/or severe viral gastroenteritis), pre-COVID health profiles (clinical indications of abnormal glucose, high BMI, triglycerides levels, pre-existing autoimmune disease (e.g., celiac disease)).

Sensitivity analysis to assess detection bias. To assess **the potential for detection** bias, Centers will plot the
dates of all RT-PCR tests (negative or positive) or initial SARS-CoV-2 diagnosis codes and diabetes diagnoses for people with incident diabetes. Kaplan Meier survival curves will also be generated for each cohort to visualize time-to-incident diabetes diagnosis. Alternate cut-off times for post-COVID diabetes will be examined.

Sensitivity analysis to assess competing risks. To examine the potential impact of competing risks on study results, a subset of sites with linkage to the national death index (NDI) will incorporate information on all-cause mortality as a potential competing event.

Exploratory analysis to stratify by additional markers of pre-COVID clinical status. Associations between SARS CoV-2 infection and new-onset diabetes may be affected by pre-COVID metabolic health and comorbidities. Models generated in Aim 3 will be revised to explore whether risk of subsequent diabetes differs according to different pre-COVID health profiles. Sites will serially stratify on the pre-pandemic high BMI, high BP, and abnormal glucose to explore the plausibility of effect modification of COVID-19-diabetes associations across strata. Variables representing pre-COVID health profiles will be derived from information *before* the SARS CoV-2 pandemic (between January 1, 2015 and December 31, 2019). Cut points for defining profiles (e.g., "high BMI") will be determined based on pre-pandemic variable distributions and clinical literature. P-values from stratified analyses will be adjusted for multiple testing. This exploratory analysis is optional.

Limitations. There are several limitations associated with the proposed analysis. First, documentation of SARS-CoV-2 exposure in the EHR is likely incomplete, especially for individuals who did not seek health care within a given health system for a SARS-CoV-2 infection. Unequal access to and uptake of SARS-CoV-2 testing may differentially impact ascertainment of COVID-19 status for individuals with and without a subsequent diabetes diagnosis. Likewise, diabetes diagnoses may be correlated with health care encounters, including encounters associated with COVID-19. While we will attempt to mitigate these potential sources of ascertainment bias by conducting an alternative, test-negative design (limited only to those who received a PCR test for SARS-CoV-2), the test-negative analysis may be impacted by "collider bias", stemming from conditioning on testing status. Residual confounding by other sociodemographic characteristics or health factors may also impact estimates. Patterns of missing data with regard to covariates may introduce selection bias, and correlates of missing data may vary across sites. Meta-analysis approaches may be subject to bias, particularly when outcomes are rare.[24]

Data Sharing and Dissemination. Analyses will be conducted locally, and summary data will be shared with the DiCAYA CoC for meta-analysis and a Network-wide manuscript. DiCAYA Network sites will sign data use agreements to share all cases that meet the working computable phenotype for diabetes (study outcome) between June 1, 2020 and December 31, 2022 and a random 5% sample of patients who meet the inclusion criteria but do not meet the computable phenotype for diabetes with the NYU CoC for data quality assurance, as well as person-time information on all study participants for Kaplan Meier curve generation, accordingly. The DiCAYA network Publications & Presentations Committee will guide authorship decisions on supplement manuscripts.

Timeline.

Below is a timeline from the original proposal that has been revised to reflect the ordering and realistic timeframe for each activity.

Table 6. Study Timeline	Quarter 1 Sept 30-Dec 31 2022	Quarter 2 Jan 1 – March 31, 2023	Quarter 3 Apr 1, June 30, 2023	Quarter 4 July 1 – Sept 30, 2023
IRB submission				
Data use agreement modifications signed				
Protocol and comparison group definitions finalized				
Aim 1. Exposure measure refinement & harmonization				
Primary propensity score variables identified & finalized; pilot propensity score models estimated				
Aim 2. Computable phenotype adaptation and characterization of diabetes onset time distributions				
Aim 3. Estimate associations between SARS CoV-2 infection and new diabetes diagnosis risk.				
Knowledge dissemination				
Legend: [] = administrative task; [] = research task; [] = dissemination task				

References

1. Fadini GP, Morieri ML, Longato E, Avogaro A. Prevalence and impact of diabetes among people infected with SARS-CoV-2. *J Endocrinol Invest*. 2020;43(6):867-869. doi:10.1007/s40618-020-01236-2

2. Zhu L, She ZG, Cheng X, et al. Association of Blood Glucose Control and Outcomes in Patients with COVID-19 and Pre-existing Type 2 Diabetes. *Cell Metab*. 2020;31(6):1068-1077.e3. doi:10.1016/j.cmet.2020.04.021

3. Pal R, Bhadada SK. COVID-19 and diabetes mellitus: An unholy interaction of two pandemics. *Diabetes Metab Syndr*. 2020;14(4):513-517. doi:10.1016/j.dsx.2020.04.049

4. Bode B, Garrett V, Messler J, et al. Glycemic Characteristics and Clinical Outcomes of COVID-19 Patients Hospitalized in the United States. *Journal of Diabetes Science and Technology*. 2020;14(4):813. doi:10.1177/1932296820924469

5. Oliveira EA, Mak RH, Colosimo EA, et al. Risk factors for COVID-19-related mortality in hospitalized children and adolescents with diabetes mellitus: An observational retrospective cohort study. *Pediatr Diabetes*. 2022;23(6):763-772. doi:10.1111/pedi.13335

6. Nagl K, Waldhör T, Hofer SE, et al. Alarming Increase of Ketoacidosis Prevalence at Type 1 Diabetes-Onset in Austria—Results From a Nationwide Registry. *Front Pediatr*. 2022;10:820156. doi:10.3389/fped.2022.820156

7. McCluskey CK, Zee-Cheng JE, Klein MJ, et al. The Temporal Relationship Between Local School Closure and Increased Incidence of Pediatric Diabetic Ketoacidosis. *Front Pediatr*. 2022;10:812265. doi:10.3389/fped.2022.812265

8. Alfayez OM, Aldmasi KS, Alruwais NH, et al. Incidence of Diabetic Ketoacidosis Among Pediatrics With Type 1 Diabetes Prior to and During COVID-19 Pandemic: A Meta-Analysis of Observational Studies. *Front Endocrinol (Lausanne)*. 2022;13:856958. doi:10.3389/fendo.2022.856958

9. Eisenbarth GS. Type I diabetes mellitus. A chronic autoimmune disease. *N Engl J Med*. 1986;314(21):1360-1368. doi:10.1056/NEJM198605223142106

10. Filippi CM, von Herrath MG. Viral trigger for type 1 diabetes: pros and cons. *Diabetes*. 2008;57(11):2863-2871. doi:10.2337/db07-1023

11. Chen S, de Craen AJM, Raz Y, et al. Cytomegalovirus seropositivity is associated with glucose regulation in the oldest old. Results from the Leiden 85-plus Study. *Immun Ageing*. 2012;9(1):18. doi:10.1186/1742-4933-9-18

12. Steenblock C, Schwarz PEH, Ludwig B, et al. COVID-19 and metabolic disease: mechanisms and clinical management. *Lancet Diabetes Endocrinol*. 2021;9(11):786-798. doi:10.1016/S2213-8587(21)00244-8

13. Khunti K, Del Prato S, Mathieu C, Kahn SE, Gabbay RA, Buse JB. COVID-19, Hyperglycemia, and New-Onset Diabetes. *Diabetes Care*. 2021;44(12):2645-2655. doi:10.2337/dc21-1318

14. Xie Y, Al-Aly Z. Risks and burdens of incident diabetes in long COVID: a cohort study. *The Lancet Diabetes & Endocrinology*. 2022;10(5):311-321. doi:10.1016/S2213-8587(22)00044-4

15. Montefusco L, Ben Nasr M, D'Addio F, et al. Acute and long-term disruption of glycometabolic control after SARS-CoV-2 infection. *Nat Metab*. 2021;3(6):774-785. doi:10.1038/s42255-021-00407-6

16. Accili D. Can COVID-19 cause diabetes? *Nat Metab*. 2021;3(2):123-125. doi:10.1038/s42255-020-00339-7

17. Rubino F, Amiel SA, Zimmet P, et al. New-Onset Diabetes in Covid-19. *N Engl J Med*. 2020;383(8):789-790. doi:10.1056/NEJMc2018688

18. McKeigue PM, McGurnaghan S, Blackbourn L, et al. Relation of Incident Type 1 Diabetes to Recent COVID-19 Infection: Cohort Study Using e-Health Record Linkage in Scotland. *Diabetes Care*. Published online July 12, 2022:dc220385. doi:10.2337/dc22-0385

19. Kendall EK, Olaker VR, Kaelber DC, Xu R, Davis PB. Association of SARS-CoV-2 Infection With New-Onset Type 1 Diabetes Among Pediatric Patients From 2020 to 2021. *JAMA Network Open*. 2022;5(9):e2233014. doi:10.1001/jamanetworkopen.2022.33014

20. Barrett CE. Risk for Newly Diagnosed Diabetes 30 Days After SARS-CoV-2 Infection Among Persons Aged 18 Years — United States, March 1, 2020–June 28, 2021. *MMWR Morb Mortal Wkly Rep*. 2022;71. doi:10.15585/mmwr.mm7102e2

21. Harding JL, Oviedo SA, Ali MK, et al. The bidirectional association between diabetes and long-COVID-19 – A systematic review. *Diabetes Res Clin Pract*. 2023;195:110202. doi:10.1016/j.diabres.2022.110202

22. Simon TD, Haaland W, Hawley K, Lambka K, Mangione-Smith R. Development and Validation of the Pediatric Medical Complexity Algorithm (PMCA) Version 3.0. *Acad Pediatr*. 2018;18(5):577-580. doi:10.1016/j.acap.2018.02.010

23. McAlexander TP, Algur Y, Schwartz BS, et al. Categorizing community type for epidemiologic evaluation of community factors and chronic disease across the United States. *Soc Sci Humanit Open*. 2022;5(1):100250. doi:10.1016/j.ssaho.2022.100250

24. Schuemie MJ, Chen Y, Madigan D, Suchard MA. Combining cox regressions across a heterogeneous distributed research network facing small and zero counts. *Stat Methods Med Res*. 2022;31(3):438-450. doi:10.1177/09622802211060518

Appendix A SEARCH 4 Registry and Cohort Protocol 9/30/2015-9/29/2021

For the SEARCH component of SEARCH-DiCAYA- Beginning in 2022 when SEARCH enters the unfunded period of the study, the SEARCH Coordinating Center at Wake Forest University will cease operations as the site organizing the study sites, housing sharable data, and organizing conference calls. The SEARCH Colorado site will take over these activities and will pursue project-specific data use and transfer agreements for any data shared from our institution to other groups. The Colorado site has housed all SEARCH study biological samples since 2020 under COMIRB 20-0136 SEARCH Consortium Stored Sample Biorepository.

1. Background

Diabetes is the third most prevalent severe chronic disease of childhood (1), and a leading cause of nephropathy, retinopathy, neuropathy, and cardiovascular disease (CVD) later in life. Although there is some evidence that rates of mortality, renal failure, and neuropathy have declined in young adults with youth-onset T1D diagnosed between the 1950s and 1980s (2), data from more contemporary cohorts are scarce. In addition, clinical care for childhood diabetes has evolved, now encompassing new insulin types and delivery systems, and new systems for monitoring glycemic excursions. Concurrently, the epidemiology of diabetes has evolved. The incidence rates of T1D have increased around the world (3, 4) and we have learned from the SEARCH for Diabetes in Youth Study that substantial proportions of adolescent minority youth now have T1D (5). Within the last two decades pediatric T2D has gone from infrequent to 15% of all diagnoses of diabetes in youth (6). Trends in the prevalence and incidence of T1 and T2D in young people are changing. Worldwide, from 1990 to 2008, the incidence of T1D increased by 2.8-4% per year (7), similar to that observed in SEARCH (8). Moreover, SEARCH demonstrated an increased prevalence of T1D between 2001 and 2009 (9). On the other hand, a recent report from Finland, with the world's highest incidence, suggested that the increase in incidence from 2005-2011 has stabilized (10). Regarding T2D, although few longitudinal studies have been conducted, there is evidence that the increase in T2D in youth stems from the increased frequency of obesity in pediatric populations (11). Interestingly, data from SEARCH suggest that prevalence of T2D may not be increasing equally across race/ethnic groups (9). Thus, there is much to be gained in studying the continued trends in incidence and prevalence of T1 and T2D.

2. Objectives

This is the fourth phase of the ongoing SEARCH for Diabetes in Youth Study. SEARCH phases 1-3 were conducted in 2000-2005, 2005-2010, and 2010-2015, respectively, and included both a registry component and a cohort component. Study methods (12) and highlights of SEARCH study findings (13) have been published and protocols are available online (www.searchfordiabetes.org). Unlike SEARCH phases 1-3, SEARCH 4 is supported by two separate grants from different funding agencies, one for the Registry Study (CDC) and one for the Cohort Study (NIH/NIDDK).

2.1 REGISTRY STUDY OBJECTIVES

In response to RFA-DP-15-002, and with funds awarded by the CDC with contribution from the NIH/NIDDK, SEARCH 4 will continue to ascertain newly diagnosed incident diabetes cases throughout the study period and one additional prevalent cohort (index year 2017) for youth age < 20 years across five geographically dispersed study centers that encompass the racial/ethnic diversity of the United States. Surveillance is framed as a tiered approach, starting with the most broad based and cost efficient approach at the highest tier (tier 1) and becoming the most focused in tier 3, optimizing use of electronic health data.

Aim 1: TIER 1 SURVEILLANCE - To ascertain prevalent diabetes cases in calendar year 2017 among youth age < 20 years at diagnosis. Research Question 1.1 What is the prevalence of diabetes in 2017, overall and by age, sex, race/ethnicity, and diabetes type? Research Question 1.2 What are the temporal trends in T1D and T2D prevalence over the three prevalent cohorts (2001, 2009, and 2017) and how do trends differ by race/ethnicity, age, and sex?

Aim 2: TIER 2 SURVEILLANCE - To continue to ascertain newly diagnosed (incident 2013-2020) diabetes cases in youth age < 20 years. Research Question 2.1 What are the temporal trends in T1D and T2D incidence since 2002 in US youth and how do trends differ by race/ethnicity, age, and sex?

Aim 3: TIER 3 SURVEILLANCE - To further determine agreement between the etiological classification of diabetes type using biochemical markers and provider assessment, to describe selected clinical characteristics at diagnosis, and to establish an infrastructure that facilitates the development of more detailed ancillary studies by storing biological samples and preserving contact with potential study participants. Data is extracted from EHRs in all incident years and an in-person visit is planned for incident cohort year 2016, using a strategic sampling plan to minimize cost. Research Question 3.1 Is the proportion of youth with provider diagnosed T1D or T2D who have biochemical evidence of these respective diagnoses consistent over time? Evidence is based on diabetes etiologic types previously established and employed by SEARCH using diabetes autoantibodies (DAA) and the insulin sensitivity (IS) score. Research Question 3.2 Has the prevalence of DKA near the time of diagnosis decreased over time for youth with T1D or T2D?

Aim 4: OPERATIONAL EFFICIENCY - To optimize efficiency of SEARCH surveillance activities through targeted Development and Validation (D&V) Projects designed to utilize electronic health data to operationalize each of the three tiers of surveillance to the extent possible. Methods employ electronic algorithms and text mining/natural language processing with validation, incorporating data from administrative records, medical records including provider notes, pharmacy, and laboratory data. We will then evaluate these approaches with a goal of identifying a model for targeted expansion of the SEARCH Registry to non-SEARCH sites.

2.2 COHORT STUDY OBJECTIVES

In response to RFA DK-14-508, and with funds provided by Special Statuary Funds for T1 Diabetes Research, SEARCH 4 will continue to follow selected incident cohorts from the SEARCH registry. Incident cohorts of youth from 2002-2006, 2008 and 2012 were asked to participate in a baseline research visit where history, demographics, health-care related variables, clinical information and factors essential for the etiologic classification of diabetes type (diabetes related-autoantibodies and markers of insulin sensitivity) (14, 15) were collected near diagnosis. Participants were asked to return at 1, 2, and 5 years from baseline for repeated measures in SEARCH phase 1 and 2. During SEARCH 3 (2010-2015), individuals who had participated in a baseline visit with at least five years duration were invited to participate in a cohort visit. At the close of SEARCH 3, 2780 individuals participated for a final response rate of 72% among eligible individuals. The current protocol (SEARCH 4 Cohort Study) will follow a subset of this cohort (as well as a subset of participants who completed a 2012 Registry In Person Visit) with another assessment to further assess risk factors, acute and chronic complications, as well as QOL-related outcomes and add measures of cardiac structure and function, neurocognitive outcomes, and social functioning and stress. We will also continue to assess mortality and causes of death.

Aim 1: Establish, compare and contrast the burden (prevalence, incidence, progression and clustering) of acute and chronic complications of diabetes, and explore the responsible risk factors and pathways among youth and young adults with T1D and T2D. We will measure key outcomes, including: retinopathy, nephropathy, cardiac autonomic (CAN) and peripheral neuropathy, arterial stiffness, cardiac damage, neurocognitive outcomes, as well as acute complications (hypoglycemia, diabetic ketoacidosis -DKA). We will explore a variety of risk factors and pathways, including: metabolic; inflammatory; vascular; behavioral; socio- economic; psycho-social and health-care factors. We hypothesize that: **1.1**: Youth with T2D have higher prevalence, incidence, faster rate of progression and different patterns of clustering of chronic complications, but lower burden of acute complications than youth with T1D, independent of age, sex, diabetes duration and race/ethnicity; **1.2:** The risk factor patterns associated with these outcomes are different in T2D vs. T1D.

Aim 2: Explore, compare and contrast processes of care (including barriers to care and quality of care- QOC) and their influence on QOL among youth with T1D and T2D, as they transition from pediatric to adult care. Measures to assess barriers include: consistent health insurance, out of pocket costs, continuity of care, employment, completion of education, finances, stressors from independence (school, work, marriage, children), social support, depression and neurocognitive factors. QOC variables include: frequency of visits with diabetes provider and receipt of screening for retinopathy, nephrology, neuropathy, foot exams, blood pressure and A1c. We hypothesize that**: 2.1:** Compared to youth with T1D, youth with T2D will a) have more and different barriers to

care; b) benefit less from emerging treatment technologies; c) have worsening QOC and QOL as they transition from pediatric to adult care.

Aim 3: Conduct surveillance of mortality including cause of death in the SEARCH cohort. We hypothesize that**: 3 . 1 :** The frequency and causes of mortality in patients with youth-onset diabetes are different than among non-diabetic, age, sex and race/ethnicity comparable persons; **3.2**: Youth and young adults with T2D have higher mortality and different causes of death than youth with T1D, independent of age, sex, diabetes duration and race/ethnicity.

Aim 4: Maintain, supplement and promote access to the SEARCH Cohort repository for biological specimens to conduct scientifically and logistically appropriate ancillary studies.

3. Study Population

3.1. STUDY SITES

The five clinical centers that participated in SEARCH 3 will continue their participation in SEARCH 4. These sites are based in Ohio, Colorado, Washington, South Carolina, and California. Four SEARCH centers (Ohio, Colorado, Washington, and South Carolina) are geographically based - that is, newly diagnosed diabetes cases are identified from a geographically defined population. One SEARCH center (California) is membership-based - that is, newly diagnosed diabetes cases are identified from the membership of the participating health plan. Each of the five centers participates in both the Registry and the Cohort Studies.

1. Ohio - Cases ascertained from Cincinnati and the 8 surrounding counties; oversight, recruitment and clinic visits provided by Children's Hospital Medical Center.
2. Colorado - Cases ascertained from the state of Colorado and members of the Navajo Indian tribe in AZ, UT, or NM residing on the Navajo Nation reservation; oversight, recruitment and clinic visits provided by University of Colorado, Denver.
3. Washington - Cases ascertained from Seattle and Tacoma and the 5 surrounding counties; oversight, recruitment and clinic visits provided by Seattle Children's Research Institute.
4. Carolinas - Cases ascertained from the state of South Carolina with oversight provided by the University of North Carolina at Chapel Hill. Sub-centers are located at three locations in SC (Charleston, Greenville, Columbia) to assist with recruitment and clinic visits.
5. California - Cases ascertained from Kaiser Permanente Southern California Health Care Plan membership (other than San Diego) with oversight, recruitment and clinical visits provided by the same.

3.2. OTHER SITES

The Coordinating Center (CC) is located at the Wake Forest School of Medicine in Winston-Salem, NC, and has served as the CC for all phases of SEARCH. The laboratories and reading centers, listed below, are supervised by and operate as subcontracts to the CC.

1. Central Laboratory- Northwest Lipid Metabolism and Diabetes Research Laboratories, University of Washington

1.2 Sample repository will move from NWLMDRL to LEAD Center at University of Colorado in Aurora, CO inSpring 2020 for stored samples. Samples collected and processed during SEARCH 4 have been or will be processed by the Central Lab in Seattle, WA.

2. Neuropathy Reading Center, University of Michigan
3. Ocular Epidemiology Reading Center, University of Wisconsin-Madison
4. Cardiovascular Reading Center, Cincinnati Children's Hospital Medical Center

3.3. STUDY POPULATION AND ELIGIBILITY

3.3.1. *The SEARCH Registry Study*

Over the three phases of SEARCH, investigators have registered more than 25,000 cases of youth with diabetes, including completed incident cohorts from 2002-2012, prevalent cohorts in 2001 and 2009, and ongoing efforts for registration of incident 2013-2015 cohorts. During SEARCH Phase 4, incident 2013-2017 cohorts will be completed, and incident 2018-2020 will be initiated but not completed (Table 1).

Table 1. Surveillance activities during SEARCH Phase 4

Phase 4 Period	Case ascertainment of youth diagnosed in:	Incident year to be closed (30 months after the end of the incident year)++:
Yr1: Oct 2015-Sept 2016	2013, 2014, 2015, 2016*	2013
Yr2: Oct 2016-Sept 2017	2014, 2015, 2016*, 2017	2014
Yr3: Oct 2017-Sept 2018	2015, 2016*, 2017, 2018+	2015
Yr4: Oct 2018-Sept 2019	2016*, 2017, 2018+, 2019+	2016
Yr5: Oct 2019-Sept 2020	2017, 2018+, 2019+, 2020+	2017

+ Registration for incident years 2018, 2019, and 2020 will not be completed during SEARCH 4. We will begin registering these cases in anticipation of future funding to fully register these incident years. Note that incident years 2013- 2015 initially began registration during SEARCH 3.
++ Beginning with incident year 2018, registration will be closed 20 months after the end of the incident year.
*In-person-visits (IPV) will be conducted on 2016 incident cases. Yr=Year.

Registry Aims 1 and 2. Centers in SEARCH Phase 4 will continue to conduct population-based ascertainment of cases of diabetes in youth less than 20 years of age for incident years 2013 through 2020, using methods consistent with those employed in SEARCH

Phases 1-3. Prevalent cases will be obtained in index year 2017. Briefly, cases are ascertained primarily though networks of pediatric endocrinologists, with pediatric diabetes databases, electronic health records from participating inpatient and outpatient settings, hospitals, and other health care organizations being queried to identify the remainder of the cases. Cases will be validated based on physician reports, medical records reviews or self-reports of a physician diagnosis of diabetes based on an established set of criteria. Further eligibility is defined by the following: 1) children/youth who, in addition to having an onset of physician-diagnosed diabetes in the index year, are also are < 20 years of age on December 31 of the index year; 2) are resident of the population defined for geographically-based centers at any time during the index year, or a member of the participating health plan for the membership-based center at diagnosis, and 3) are not active duty military personnel or institutionalized. Young women who develop gestational diabetes mellitus (GDM) but who are not diagnosed with diabetes when not pregnant are not eligible. Sites are provided a 30 month window after the close of the incident year to identify all potential cases. For example, for incident year 2016, the window closes 6/30/2019 (Table 1). A total of 13,440 incident cases are expected to be registered during SEARCH Phase 4 (Table 2). Beginning with incident year 2018, registration will be closed 20 months after the end of the incident year.

The prevalence study for 2017 will attempt to identify and validate all unique, eligible cases of diabetes in youth less than 20 years who are residing in or are members of the SEARCH geographic areas and health plans in 2017. Previous prevalence studies have been conducted in SEARCH in 2001 and 2009. A total of 1004 new prevalent cases not previously identified through the incidence study are expected to be registered (Table 2). Completeness of case ascertainment will continue to be monitored via capture-recapture analyses, as described in detail on the SEARCH website (16).

Table 2: Estimated Number of Registered Cases (Incident and Prevalent) and IPV, Overall and By Site, SEARCH 4

	Carolinas	Ohio	Colorado	California	Washington	Total
2013 Incident*	378	179	422	265	284	1511
2014 Incident*	389	184	435	273	292	1557
2015 Incident*	401	189	448	281	301	1603
2016 Incident	413	195	461	290	310	1652
2017 Incident	425	201	475	299	320	1701
2018 Incident	438	207	489	307	329	1752
2019 Incident	451	213	504	317	339	1805
2020 Incident	465	220	519	326	349	1859
Total Incident	**3360**	**1588**	**3753**	**2358**	**2524**	**13,440**

2017 Prevalent**	57	112	86	508	241	1004
Total Cases (I + P)	**3417**	**1700**	**3839**	**2866**	**2765**	**14,444**
Total IPV* **	**207**	**81**	**220**	**171**	**153**	**832**

* Total number of cases we expect to register for these incident years, including those registered under the SEARCH 3 protocol. **Excludes incident 2017 cases, and all previously registered incident and prevalent cases included in the 2017 prevalent sample. *** IPV for 2016 incident cases.

The calculation of incidence and prevalence rates require information on the population at risk. Race-bridged post-censal estimates of the July 1 resident US population, released yearly by the National Center for Health Statistics, are used as the denominators for the geographic sites. Each file contains population estimates for each US county by single year of age, bridged-race, sex, and Hispanic origin. Active duty military are excluded. The membership site (California) uses July 1 health plan enrollment data by single year of age and sex as the denominator. Addresses for each of the members are geocoded and census block level data are used as a source of race/ethnicity (17). The Indian Health Service user population for eligible service units on the Navajo Nation, defined as persons age < 20 years with one or more visits in the past 3 years (including the index year) is used to estimate denominators for this Colorado sub-site.

Registry Aim 3. A sample of cases diagnosed in 2016 will be invited for an in-person visit (IPV). Cases eligible for the IPV will include all cases diagnosed during 2016 who are of minority race/ethnicity, those with a provider diagnosis of T2D, and 25% of non-Hispanic white youth with a provider diagnosis of T1D, randomly selected for invitation. This sampling plan will yield approximately 832 IPVs (Table 2).

3.3.2. *The SEARCH Cohort Study*

Cohort Study Follow-Up (Cohort Aims 1 and 2). A subset of SEARCH 3 Cohort (C_1) and SEARCH 3 Registry (R_1) participants will be invited for a SEARCH 4 in-person visit. The eligible group will include all SEARCH 3 (C_1 and R_1) participants with T2D, all minority youth with T1D, and a random sample of NHW youth with T1D. The Coordinating Center (CC) will provide a list of randomly selected NHW youth with T1D to be invited for participation such that all participants will be 10 years or older, have at least 3 years of time elapsed since their SEARCH 3 (C_1 or R_1) visit and have at least 5 years of duration of diabetes at the time of their planned SEARCH 4 IPV. NHW sampling is performed since based on the limited available budget it was determined that there was minimal gain in statistical power to invite all T1 NHW youth for a return visit and that all proposed analyses could be addressed with the random sample of NHW T1. Table 3 shows the total number of participants expected to complete a SEARCH 4 Cohort visit (N~1,846) based on the proposed sampling. These estimates are based on an expected 75% response rate. In addition, the SEARCH 4 IPV will include a sample of 500 participants to be identified by the CC to have cardiac echocardiogram measurements

taken. This sample will include 250 T1D and 250 T2D with representation from all five clinical sites and have racial/ethnic diversity.

The remainder of SEARCH 3 (C_1) participants will form the survey-only group with no IPV in SEARCH 4. The survey-only option will also be offered to individuals who are eligible but refuse participation in the IPV. The survey-only group will be asked to complete questionnaires by mail, phone or internet. Survey data will be combined from the IPV and survey-only participants (at least 2,546) to address Aim 2.

Table 3. Number of Expected participants for each component of the SEARCH 4 Cohort Study, by Clinical Site

Site	In-Person Visit	Echocardiogram	(Estimated) Survey Only
Carolinas	434	140	150
Ohio	298	130	150
Colorado	504	130	225
California	352	0	25
Washington	258	50	150
Total	1846	450	700

Mortality follow up (Cohort Aim 3): All incident cases identified by the Registry study during calendar years 2002-2015 will be included in the mortality follow-up through 12/31/17 using the National Death Index (NDI) (18). This is the second mortality assessment, with the initial one done for incident cases identified during 2002-2008 and followed through 12/31/10. Mortality status will be obtained by matching with the NDI as soon as the NDI has complete data for 2017, usually ~18 months from the close of the time period. Conservatively, we estimate that there will be 89 additional deaths for a total of 130, using mortality rates from the prior period. This will allow us to examine cause-specific deaths in selected subgroups.

In order to reduce the amount of PHI sent offsite to the National Death Index, the SEARCH Colorado site will first identify a list of mortality surveillance eligible cases with unknown mortality statuses and send that list with the minimum PHI necessary to determine a match to the Health Data Compass team using encrypted and password protected secure upload in RedCap. Only cases identified with unknown mortality statuses will be sent to the National Death Index for matching.

3.4. INFORMED CONSENT

Consent for the SEARCH Study is handled through three important mechanisms under the supervision of local IRB(s). Individual differences exist based on requirements of local IRBs. In general: 1) initial data for the Registry Study is collected (without participant contact) under a HIPAA waiver; 2) completion of surveys (either by mail or web-based/online) is covered by a waiver of documentation of consent [aka, implied consent] according to local IRB requirements in the Registry and Cohort Studies; and 3) written informed consent is obtained prior to all IPVs in the Registry and Cohort Studies.

As in previous phases of SEARCH, the initial data collection in the Registry Study (case ascertainment) is covered by a HIPAA waiver. That is, identification of all new cases of diabetes in a defined geographic area or health plan does not require that registered cases provide written or implied informed consent; HIPAA requirements are waived.

Mailed and/or web-based online surveys are utilized in both the Registry Study and Cohort Study. In this case, consent is implied with completion of the surveys. In the Registry Study, potentially eligible cases are mailed (or emailed with internet link) an introductory letter that gives a brief description of the research study along with the Initial Participant Survey (IPS). For individuals who are less than 18 years of age, the introductory letter is mailed to a parent or guardian. If the completed IPS is not returned and the participant does not refuse after receiving the introductory letter, a member of the local research team may call the individual or the parent to complete the IPS. Again, consent requirements for completion of the IPS are governed by the local IRB. In the Cohort Study, a subset will be asked to participate in the survey-only group, for which surveys will be mailed (or emailed with internet link) for completion at home.

Written informed consent is obtained for all individuals/parents who agree to participate in the Registry Study IPV as well as for the Cohort Study IPV in accordance with local IRB requirements. If the participant is less than 18 years of age, the parent or guardian must give written informed consent prior to the initiation of any study procedures or data collection, according to the requirements of the local IRB. Written assent of participants who are less than 18 years of age is also governed by the requirements of the local IRB. If the participant is 18 years of age or older, the participant must give written informed consent. Copies of completed consent forms are maintained in the participant's local research record.

There are three optional components to the written informed consent: storage of serum/plasma/urine and DNA/miRNA; transfer of data and samples to the NIDDK Repository; and sharing of data and genetic information with dbGaP (**d**ata**b**ase of **G**enotypes **a**nd **P**henotypes). In each case, participants or their parent must indicate in writing whether or not they are providing consent for these optional components. The NIDDK Central Repository is a research resource supported by the National Institutes of Health. At the end of the SEARCH, de-identified research data and samples of blood and urine will be provided to the Repository

for participants who have consented to this component. For all optional components, participants may choose to participate in SEARCH but not provide consent to participate in these components.

3.5. RECRUITMENT & RETENTION

The SEARCH Registry Study sites continue to employ a wide variety of methods shown to be highly effective at recruiting study participants for the IPS and IPV. Recruitment strategies have included meeting the family at a medical appointment to introduce the study; mailing study brochures and other informational letters; posting study materials in clinics; enlisting the encouragement by diabetes care providers; emailing, texting, or using social media such as Facebook to contact potential participants; phoning participants to complete the surveys and/or schedule a visit; offering online surveys; and one or more reminder calls prior to the scheduled visit. Participation in the IPV is facilitated by flexible weekday appointments, as well as Saturdays, satellite clinics, and home visits. Sites offer to pair research visits with clinical appointments when possible; provide transportation and/or lodging; and generally assist participants with removing barriers to study participation. Study participants are offered remuneration that is appropriate for the length and burden of the study visit. Participants and their providers receive the clinically-relevant research laboratory test results, which may assist with their clinical care. To retain Registry Study participants for future studies and to share study progress, we utilize traditional, proven, retention strategies including: birthday cards, study newsletters, updating contact information annually, and utilizing internet-based search systems to locate individuals lost to follow-up.

Similarly, the SEARCH Cohort Study has maintained outstanding participant retention throughout its history. We continue to employ traditional, proven, retention strategies as described above. We also offer flexible study date appointments including home visits, offer assistance with transportation, mail pre-visit instructions, one or more reminder calls prior to the scheduled visit, provide acknowledgement of participation, and provide participant remunerations that are appropriate for the length and the respondent burden of the proposed study visit. Investigators and study personnel also continue to solicit the support of diabetes providers to encourage on-going study participation. Communications with providers include letters, e-mail messages, telephone calls, newsletters, individual discussions, and group presentations of study goals and preliminary results.

4. Study Measurements

For SEARCH phases 1-3, all clinical sites have operated under a common protocol. This approach is followed in SEARCH 4 Registry and Cohort Studies as well. That is, data from each site is obtained, managed, and protected according to a standard study protocol that has been developed and vetted by the Steering Committee and approved by all participating IRBs and by the NIDDK Observational Studies Monitoring Board (OSMB). Clinic sites use a standard informed consent template, modified as needed by local IRB requirements. All clinic staff are

trained and certified, operate under a single Manual of Procedures (MOP), and follow a standard set of data collection procedures. Clinic staff participate in both central and local training as needed. Clinical Center investigators and staff participate in ongoing working groups and established study committees to ensure that identical procedures are followed at each site for the purpose of recruitment, retention, and ensuring the highest quality of study data.

4.1. MEASUREMENTS - REGISTRY STUDY

Centers in the SEARCH Registry Study continue to conduct population-based ascertainment of cases of diabetes in youth < 20 years of age using methods consistent with those employed in SEARCH 1-3. This involves identification, case validation, confirmation of eligibility, deduplication, and registration of cases centrally with the SEARCH Coordinating Center. There are three aspects of data collection in the Registry Study: 1) data obtained from all potential registered cases; 2) the Initial Participant Survey (IPS); and 3) data obtained during an in-person visit (IPV) on a subset of the 2016 registered cases.

Collection of Data on all Registered Cases: A minimum amount of demographic and clinical information is needed for all cases in order to calculate population-based incidence rates and prevalence of diabetes mellitus by age, sex, diabetes type and race/ethnicity. The primary source of this information is the medical record except for race and Hispanic ethnicity, which, when obtained by self-report using the IPS, supersedes the report via medical record. Study staff abstract information from the medical record for the period from diabetes diagnosis to six months after this date to obtain the following information: 1) date of birth, 2) sex, 3) race/ethnicity, 4) diagnosis date, 5) zip code at diagnosis, 6) county and state of residence at diagnosis, 7) diabetes type at the time of diagnosis and the diabetes type reported closest to 6 months, 8) whether diabetes autoantibodies were measured up to 6 months after diagnosis [GAD/GAA, IA2/ICA512, ICA, IAA, and ZnT8], 9) height, 10) weight (closest to diagnosis), 11) whether the participant ever used insulin, 12) whether insulin was discontinued, 13) presence of acanthosis nigricans, and 14) whether DKA was noted (with dates, bicarbonate, pH, and glucose values). For potential cases not eligible for registration, minimal demographic data are maintained in order to facilitate validation and de-duplication of local cases.

Initial Participant Survey (IPS): All registered cases are invited to complete the IPS. The IPS is used to: a) verify case eligibility (e.g., residence in the year of diagnosis); b) obtain self-reported race/ethnicity and selected clinical and demographic information; and c) introduce participants to SEARCH to facilitate future studies. The IPS queries symptoms at presentation, potential secondary causes of diabetes, use of insulin and other medications, diabetes treatment history, height and weight, family structure, usual language spoken, type of health insurance, usual provider for diabetes care, highest parental education, household income, nativity of person with diabetes and their parents, and contact information. All registered cases are eligible to complete the IPS online, by mail, or by interviewer administration by telephone or in person.

In-Person Visit (IPV): A sample of registered cases diagnosed in 2016 will be invited to an IPV. The IPV enables an analysis comparing agreement between provider assigned diabetes type compared to SEARCH etiologic type in order to interpret the potential meaning of trends over time according to provider type, and to enable statistical adjustment for differences in agreement over time. The IPV, lasting approximately 60 minutes, includes collection of fasting blood (4-5 TBSP) and urine samples, a brief physical examination, and a medication inventory, all conducted under SEARCH standardized protocols and described below for the Cohort Study. Measurements made to inform dimensions of diabetes type include diabetes autoimmunity (GAD65, IA-2, and ZnT8 antibodies), and the SEARCH validated insulin sensitivity index (waist circumference, HbA1c, and triglycerides) (15). We will also measure markers of kidney function (albumin and creatinine from a first morning void, cystatin-c and serum creatinine), the latter two measures pending availability of funds. To facilitate work that requires additional funding in the future, we will store plasma, serum, DNA, and urine. Specimens are processed locally and shipped within 24 hours to the central laboratory. Diabetes autoantibodies are measured by standardized protocol and a common serum calibrator developed by an NIDDK- sponsored standardization group.

4.2. MEASUREMENTS - COHORT STUDY

The study visit for the cohort study participants is expected to take approximately four hours and includes physical measures and questionnaires. A parent/guardian is required to attend if the individual with diabetes is < 18 years old. Most of the measures obtained during the SEARCH 4 visit are the same as to those obtained in previous visits. New measures, particularly cardiac echocardiography and neurocognitive testing, are noted in Table 4, along with data obtained at C_0 (baseline visit), intermediate visits (12, 24, 60 months), and the SEARCH 3 Cohort visit (C_1).

Surveys. SEARCH has included surveys in multiple domains over time. Surveys in SEARCH 4 include:

a) Health history including pregnancy history of women;

b) Treatment including all prescribed medications, insulin regimens and glucose monitoring devices (19,20);

c) Behavioral factors including diet (21), physical activity (22), TV and computer use (23), smoking (24), and substance use (25);

d) Psychosocial factors using CES-D scales (26-28), the PROMIS and PHQ9 depression and anxiety screening tools (29), Hypoglycemia Fear Survey (HFS-C,P) (30-33), the updated Diabetes Responsibility and Conflict Scale to assess diabetes-specific family conflict (34); Stigma and Discrimination. Diabetes self-care is assessed with the Diabetes Self-Management Questionnaire (DSMQ) (35);

e) Socio-cultural factors including household and per capita income, family structure, preferred

language, migration status, parental and participant attained education, participant employment status, household food security;

f) Processes of care including type and frequency of utilizing health care providers, processes of diabetes self-management training, and recent hospitalizations (36);

g) Quality of care based on ADA guidelines for pediatric diabetes care in terms of testing frequency for HbA1c, blood pressure, lipids, urine albumin, retinopathy screening, and foot checks (37). Receipt of services is measured by self-report by parents (participant age <18) or adult participants (age ≥ 18 yrs.);

h) Quality of life using the Pediatric Quality of Life Inventory (PedsQL) (38-40) with age-specific and parental scales for participants < 18 years and validated scales for young adults 18-25 and over 26;

i) Barriers to care via items from the Consumer Assessment of Healthcare Providers and Systems survey (CAHPS 3.0) Supplemental Item Set for Children with Chronic Conditions. Additional information about continuity of health insurance, continuity of care, cost-related non-adherence and financial burden is collected using the following surveys, adapted for youth and young adults: Medical Expenditure Panel Survey (MEPS) [Agency for Healthcare Research and Quality (AHRQ)]; Perceived Financial Burden of Diabetes and Cost-related Medication Non-adherence (41);

j) Transition to adult care: Specific questions about processes of care, motivations, satisfaction with, and preparation for transition from pediatric to adult care, adapted from validated measures that have been developed to assess patients' perceptions of other kinds of care transition, such as the Care Transition Measure (42). We will also measure care transition planning by adapting items from the National Survey of Children with Special Health Care Needs.

Physical exam. Standardized anthropometry methods include height, weight, waist circumference (using NHANES and WHO protocols); systolic and diastolic blood pressure; and evaluation for acanthosis nigricans.

Laboratory parameters. Fasting blood (4-5 TBSP) and first morning urine are collected following standard protocols. Blood and urine laboratory parameters continue to be measured using established protocols at the Northwest Lipid Research Laboratory. Samples are shipped from clinical centers to the central laboratory. Results are sent from the laboratory to the CC through established secure protocols. In the spring of 2020, store samples remaining for future research use will move from Northwest Lipid Research Laboratory to the LEAD Center SEARCH Biorepositoty at the University of Colorado Denver Campus in Aurora, Colorado.

Cardiac echocardiography. Measures of cardiac structure and function are obtained using cardiac echocardiography in a subgroup of Cohort Study participants. Measures include two dimensional (2-D) directed M-mode echo images to determine left ventricular mass (LVM), left atrial size and relative wall thickness, as well as shortening fraction, LV strain and diastolic function. The primary outcome measure is LVM determined by 2-D guided M-mode echo at end diastole (43, 44) using the autopsy corrected equation of Devereux (45). Echocardiograms are read on a Digiview instrument and strain is read on a Tomtec instrument. Digital images recorded on CDs identified only by participant ID number are sent to Cardiovascular Reading Center.

Retinal Photography. We will continue obtaining retinal images using Canon CR-1 Mark II fundus cameras. Consistent with NHANES protocol (46), two 45-degree images are taken of each eye: one centered on the optic nerve and the other on the fovea. The Ocular Epidemiology Reading Center at the University of Wisconsin-Madison (47) will grade the images for presence and severity of diabetic retinopathy (DR), macular edema and will make measurements of retinal vessel calibers. After grading the retinal images from the 2nd retinal visit, a separate longitudinal review will be conducted to confirm progression/regression status of diabetic retinopathy or macular edema severity.

Table 4. Data Collected on Cohort Study Participants

Variables	Baseline Visit (C_0)	12, 24, 60 months	SEARCH 3 Visit (C_1)	SEARCH 4 Visit (C_2)
Surveys:				
Demographics: Sex, Race/ethnicity, Parental age	X			
Employment, education: parent or youth > 18 years		X	X	X
Medical Record: Diabetes type, date of diagnosis	X			
Health History: Birth date &weight, age at onset	X			
Pubertal status, co-morbidities; family history	X	X	X	X
Pregnancy outcomes in females				X
Medication: Diabetes & related conditions	X	X	X	X
Behavioral: Diet, physical activity, alcohol use	X	X	X	X
Marijuana, other substance use				X
Processes of care/quality of care			X	X
Health care costs			X	X
Psychosocial: Depression (CES-D)	X	X	X	X
Family conflict; fear of hypoglycemia			X	X
Transitions of care			X	X
Food security and assistance				X
Stressors; work ability index; stigma/discrimination				X
Physical exam: BMI, waist , blood pressure, acanthosis	X	X	X	X
Laboratory measures (blood):				
Autoantibodies	X	X	X	
Fasting glucose, cystatin C, serum creatinine, fasting C-peptide, lipid profile, inflammatory markers (CRP, IL6), A1c, AGE (CML), DNA, miRNA extraction	X	X	X	X

URINE: albumin, creatinine (spot)	X	X	X	
URINE: albumin, creatinine (first morning void)			X	X
Stored Samples: DNA, miRNA, serum, plasma, urine	X	X	X	X
Outcome(s):				
Cardiovascular: Arterial stiffness (PWV, AiX)	X		X	X
Cardiac echocardiography: LV mass, systolic & diastolic function				X
Neuropathy: heart rate variability; peripheral neuropathy		X (pilot)	X	X
Retinopathy Retinal photos, vessel caliber		X (pilot)	X	X
Nephropathy: Albuminuria	X	X	X	X
Cystatin C		X	X	X
Neurocognitive tests: NIH Toolbox.				X
Acute complications: DKA, hypoglycemia	X	X	X	X
Quality of life (Peds QL3.2 Diabetes Module)	X	X	X	X
Mortality surveillance (NDI)	X	X	X	X

Measures of Kidney Function

Urine albumin:creatinine ratio: We will collect first morning void (FMV) urine samples for storage and calculation of urine albumin:creatinine ratio (UACR). In previous phases we had collected random urine samples (with the addition of the FMV at SEARCH 3). A pilot study was introduced in April 2019 to test the feasibility of participant mailing of repeat urine samples at the Colorado and Carolinas sites.

Estimated glomerular filtration rate: Equations with the most accurate and precise estimation of glomerular filtration rate (GFR), utilize both serum creatinine and cystatin C (48, 49). Both tests have been measured in SEARCH 1-3 and continue to be measured in SEARCH 4. Different equations are currently used in children versus adults, and on the expected range of GFR (hyperfiltration versus normal GFR versus low GFR) (48-51). The natural history of eGFR in diabetic kidney disease can be heterogeneous and so we will also investigate the optimal equations for use in children versus adults and at different spectrums of GFR.

Measures of Neuropathy

Peripheral Neuropathy: The Michigan Neuropathy Screening Instrument (MNSI) will be used to screen for the presence of diabetic neuropathy. It consists of 15 self-administered questions on foot sensation including pain, numbness and temperature sensitivity. The second part of the MNSI is a brief physical examination involving 1) inspection of the feet for deformities, dry skin, hair or nail abnormalities, callous or infection, 2) semi-quantitative assessment of vibration sensation at the dorsum of the great toe, 3) grading of ankle reflexes and 4) monofilament testing. Patients screening positive on the clinical portion of the MNSI (greater than 2 points on a 10 point scale) are considered neuropathic.

Cardiac Autonomic Neuropathy: Heart rate variability (HRV) analysis allows us to assess the autonomic nervous system by examining sympathetic balance, which raises heart rate and

blood pressure and causes vasoconstriction, and the parasympathetic balance which has opposite effects (52,53). Assessments use a SphygmoCor SCOR-CPV device (AtCor Medical, Australia) as performed previously (54).

Arterial stiffness. Pulse wave velocity (PWV) is measured using the SphygmoCor (55). The average of 3 ECG R-wave gated arterial waveforms are recorded from the carotid and then the femoral arteries. Augmentation index (Aix*)* is measured with the same device (56).

Neurocognitive tests. Neurocognitive measures are computer administered utilizing the NIH Toolbox (57); domains include attention, verbal skill, working memory, mental flexibility, episodic memory, speed of processing, and response inhibition. These areas were chosen to reflect both more generalized (depressed psychomotor speed) and distinct areas of deficit (memory, attention, and mental flexibility). Receptive language vocabulary is used as a proxy for educational attainment/premorbid functioning.

Acute complications. Acute complications studied are severe hypoglycemia and diabetic ketoacidosis (DKA). Severe hypoglycemia is defined as a hypoglycemia event requiring assistance of another person (58). For DKA, occurrence is recorded as emergency department visit or hospitalization. This aligns well with prior publications on acute complications, and data frequently recorded in patient surveys and medical records (59).

Mortality Surveillance. All centers will systematically identify deaths that occur between the date of diagnosis and December 31, 2017 among youth in the 2002-2015 incident cohorts. The National Death Index will serve as the primary data source, plus individual case reports of deaths made to the study team during the course of the study. The Colorado site will augment these efforts with assistance from Health Data Compass.

5. Development and Validation Projects

The recent implementation of robust EHR systems throughout the US provides opportunities to substantially enhance the efficiency of surveillance and to pilot expansion of the SEARCH Registry beyond the currently funded sites. SEARCH 4 will attempt to optimize efficiency of SEARCH surveillance activities through targeted Development and Validation (D&V) Projects designed to utilize electronic health data to operationalize each of the three tiers of surveillance. Methods will employ electronic algorithms and text mining/natural language processing with validation, incorporating data from administrative records, medical records including provider notes, pharmacy, and laboratory data. We will then evaluate these approaches with a goal of identifying a model for targeted expansion of the SEARCH Registry to non-SEARCH sites.

Four of the five SEARCH centers are part of networks funded by Patient-Centered Outcomes Research Institute (PCORI)'s multi-institutional clinical data research networks (CDRN) formed in 2014. Three of the centers, OH, WA, and CO, are part of "A National Pediatric Learning Health System Network" (PEDSnet). The fourth clinical center, CA, is part of "Kaiser Permanente & Strategic Partners Patient Outcomes Research to Advance Learning" (PORTAL)

Network. In South Carolina, a new entity, Health Sciences South Carolina (HSSC), has been establishing a data warehouse to bring together EHR data from at least four of the six major provider systems from which cases are ascertained for the Carolina site.

This work will follow a three-step process to include development, validation and implementation. First, new approaches will be developed and initially validated through the D&V Projects in limited locations. For each approach, the established SEARCH processes for case ascertainment, case validation, and determination of diagnosis date, diabetes type, and key clinical and demographic data will be considered the "gold standard" against which new approaches will be compared. Second, approaches that meet appropriate initial validation criteria will be further refined and validated at additional SEARCH centers. Third, implementation as part of ongoing SEARCH Registry work will occur only after pre-determined metrics (e.g., sensitivity, specificity, PPV) are demonstrated for each EHR system in which the approach is to be implemented.

5.1 PROJECT #1: CASE ASCERTAINMENT BY DIABETES TYPE

The goal of D&V Project 1 is to maximize the automation of ascertainment of diabetes cases, overall and by diabetes type, by applying case identification algorithms and text analytics/natural language processing (NLP). This work is critical to Tier 1 (Prevalence) efforts. Two approaches will be employed building on previously described SEARCH work (17, 60) one based on algorithms using EHR and administrative data, the second using natural language programming (NLP) to extract and analyze text. We will attempt to replicate our previous work using EHR-based algorithms as developed in the Carolinas site to determine if these algorithms perform in a similar manner in an integrated health care system in the California site. Additionally, we will apply case identification algorithms to the PEDSnet data for the three SEARCH sites, and compare results to those using the SEARCH gold standard methods using metrics described above. Regarding the text analytics approaches, we propose to apply work as developed in the Carolinas site, including re-training of the machine learning models, to clinical notes from at least one provider from each of the five SEARCH centers, to include the three SEARCH centers that are part of PEDSnet, the CA center utilizing the KPSC data systems and Carolina working with HSSC.

5.2PROJECT #2: DETERMINATION OF DIAGNOSIS DATE

The critical information element that distinguishes Tier 1 (Prevalence) surveillance from Tier 2 (Incidence) surveillance is date of diagnosis, which generally is not available as a structured data element that can be easily extracted from the EHR. Thus, the current literature that describes various EHR-based algorithms for case identification is generally applicable only to prevalence. The goal of D&V Project # 2 is to use electronic ascertainment methods to determine diabetes diagnosis date with an expectation that at least 95% of the estimated dates will identify the correct calendar year of diagnosis. As in D&V

Project #1, two approaches will be employed: 1) use of EHR-based algorithms applied to structured data; 2) use of text analytics applied to unstructured data.

All five SEARCH centers will participate in the EHR-based algorithm work. Regarding use of text analytics, the Carolinas site will continue to refine the machine learning model with the goal of attaining at least 90% accuracy for year of diagnosis. Once optimal algorithms are established for T1 and T2D, we will expand the effort to build on the text analytic work being done for D&V Project # 1 at each of the five SEARCH centers.

5.3 PROJECT #3: AUTOMATION OF CARE AND CLINICAL DATA COLLECTION

The third project will focus on whether the collection of core and selected demographic and clinical information can be automated by directly importing information from the EHR and other clinical and administrative data systems. In addition to data obtained as part of case ascertainment by diabetes type (Project #1, Tier 1) and date of diagnosis (Project # 2, Tier 2), additional information of importance includes race and ethnicity, measurement of diabetes autoantibodies, clinical information including laboratory values related to diabetic ketoacidosis, diabetes medications, etc. The evaluation of the data capture procedures will consider both completeness as well as the accuracy of the information extracted compared to manual extraction of core data.

5.4 PROJECT #4: EXPANSION OF SURVEILLANCE TO ADDITIONAL AMERICAN INDIAN (AI) TRIBES

Since its inception in 2000, SEARCH has been conducting surveillance of youth onset diabetes in AI tribes under the direction of the Colorado site. These results indicate that AI youth have the highest incidence and prevalence of T2D of any major race/ethnic group (61, 6, 9). Unfortunately, the AI population under surveillance is the smallest of the major race/ethnic groups (~95,000 youth), and results in less than ~40 incident cases per year across all sites. The goal of this project is to develop and validate an algorithm that may be used to identify AI and possibly Alaskan Native (AN) youth with diabetes using data extracts from *existing* electronic health records (EHR) of the Indian Health Service (IHS). For the proposed pilot project, the SEARCH Colorado site will partner with the Center for AI AN Health (CAIANH), both located within the Colorado School of Public Health.

An algorithm will be developed to identify AI youth aged < 20 years with diabetes using IHS data that includes diagnostic codes, provider and service type information and dispensed medications. The algorithm will be developed from the IHS data for the Chinle and Tuba City Service Units on the Navajo Reservation and validated by comparison to the Navajo SEARCH registry (gold standard) for the same service units and will result in metrics for sensitivity, specificity, positive and negative predictive value. Next, and coordinated through CAIANH, the best algorithm developed in Phase 1 will be used to identify AI youth with

diabetes in another IHS Service Unit, from a different tribe, using the same IHS National Data Warehouse.

5.5 COST OF THE REGISTRY

Efforts to enhance efficiency are driven, in part, by potential cost savings. In SEARCH 4, we will conduct a prospective assessment in the two parallel aspects of SEARCH: conventional case ascertainment (Aims 1-3) and the D&V projects (Aim 4) in order to estimate the cost of case registration. The primary means of data collection will be the time diary in which staff members will be asked to record all SEARCH activities over a typical work week, periodically over time, with attention to infrequent tasks (e.g., those conducted monthly). The types of activities to be tracked include: managing people, clerical (mailing, logging, filing), training of staff, IT support, meetings, locating/reviewing/entering data, identification/validation/deduplication/registration, analyzing/generating reports, and local travel. For the conventional case ascertainment, diaries will be completed one week each quarter, over a period of one year. For the D&V projects, diaries will be completed more frequently, depending upon the length of the project. Actual salary and benefit rates will be applied to the time elements. A count of the number and type of cases registered during the time period will be obtained from the SEARCH registration database. Specifically, we will evaluate time and cost for the Registry as it is currently conducted, then will systematically model the incremental differences that can be attributed to approaches determined to be valid from the D&V Projects.

6. Statistical Considerations

5.6 REGISTRY STUDY - STATISTICAL CONSIDERATIONS

5.6.1 Aim 1: Detectable Differences in Prevalence

The third assessment of the prevalence of diabetes in youth is scheduled for 2017. Similar to previous work, prevalence will be expressed as the number cases with T1 or T2D per 1,000 youth pooled across all SEARCH sites. Prevalence estimates will be derived by sex, age and by race/ethnicity groups within each diabetes type. Trends in prevalence will be assessed by comparing the 2017 estimates to those observed in 2009 and 2001. Poisson regression models will be fitted to incorporate results from all 3 surveys. Standard errors associated with the estimated change in prevalence rates between any 2 time points will be computed using a 2 sided skew-corrected inverted score tests for binomial distribution. Standard error for the trends in prevalence estimates will be derived from the Poisson regression model. This model will also be used to generate adjusted prevalence where adjustment will be made for race/ethnicity, age and sex. Our power calculation suggests that we are well-powered to detect changes in prevalence by diabetes type, and across race/ethnic group within each diabetes type. For example, we will have at least 90% power to detect a rate of change of 4.1% in NHW

youth with T1D, and a rate of change of 19.1% in NHB youths with T2D for the period between 2009 and 2017.

5.6.2 Aim 2: Detectable Differences in Incidence

A similar approach will be taken to estimate the incidence rates of diabetes by type, race/ethnicity, sex and age. Incidence rates will be estimated as the number of diagnosed cases across all sites divided by the total number of individuals who are at risk across these sites. The incidence rates will be expressed in terms of the number of cases diagnosed per year per 100,000 individuals. Adjusted incidence rates will also be provided by race/ethnicity, sex and age. SEARCH 4 will add 5 additional years of incidence data taking the current time series from 12 to 17 years of data, thereby providing improved power to detect changes in incidence rate during this period for various subgroups. Based on our power calculations (see Table 5), SEARCH 4 will have 90% power to detect changes as small as 1.04% in NHW females with T1D and 2.1% in NHB females with T2D. However, we will have limited power to detect changes in API and AI youth. This is the rationale for D&V Project # 4 in which we propose an approach to develop a model for extension of the SEARCH Registry to increase inclusion of population subgroups for which our sample size is limited.

Table 5.: Detectable rate of change in incidence rate by diabetes type, sex and race/ethnicity, and power

Race	Sex	Type 1		Type 2	
		90%	80%	90%	80%
All	All	0.6	0.5	1.1	0.9
	F	0.9	0.8	1.4	1.2
	M	0.8	0.7	1.8	1.5
NHW	All	0.7	0.6	2.5	2.2
	F	1.0	0.9	3.2	2.8
	M	1.0	0.9	4.0	3.4
Hispanic	All	1.5	1.3	2.0	1.7
	F	2.1	1.8	2.6	2.2
	M	2.1	1.8	3.1	2.7
NHB	All	1.9	1.6	1.8	1.5
	F	2.7	2.3	2.1	1.8
	M	2.7	2.3	3.4	2.9
API	All	4.7	4.1	4.7	4.1
	F	7.4	6.4	6.9	6.0
	M	6.1	5.3	6.5	5.6
AI	All	8.0	6.9	4.2	3.6
	F	11.4	9.8	6.6	5.7
	M	11.2	9.6	5.3	4.6

Detecting a "leveling off" of T1D incidence in NHW Youth. The first 8 years of incidence data collected during the 2002-2009 period suggests a linear trend with a constant rate of increase of about 3% per year. With the accumulation of 5 more years of data SEARCH could be in a position to detect potential changes in incidence trends, and estimate retrospectively the incident year when the change happened. Simulation studies were performed to assess the power to correctly identify the year corresponding to the change point. The simulation study started with the data that is already available in SEARCH, which was used to fit Poisson regression models, which was then used to predict yearly incidence rates until 2018 assuming a linear trend. The model is then perturbed to mimic the effect of a change point that could occur respectively in 2012, 2013, 2014, 2015 and 2016. The perturbed model assumes that the reduction in incidence rate happened at the selected year and remained constant at the new rate in future years. This simulation process indicated that we will have ~70% power to detect a reduction of 5% in the incidence rate after 2016. It should be noted that the Finnish T1D registry study needed more than 30 years of data to be able to retrospectively identify 1988 and 2002 as the years where

changes in the incidence rate happened, with only the change point observed in 1988 being statistically significant (62). Our proposed analysis will be conducted in the second half of 2019 – after the close of the 30 months window for the incident 2016 cases.

Adjusting Results for Potential Differences in Agreement between Provider Type and Etiologic Type Over Time. Estimation of incidence trends can be affected by potential temporal changes in provider assessment of diabetes type. Such changes can lead to biased estimation of the trend. We will test for homogeneity of association between diabetes type as assessed by the provider and SEARCH etiologic type over the time period, and adjust for the difference in agreement over time as needed.

5.6.3 Aim 3: Detectable Differences in Prevalence of DKA

Assuming a significance level of 0.05, we have 80% power to detect an absolute change of 0.22% (from 30.3% to 30.1% for example) in the prevalence of DKA among T1D cases, and an absolute change of 0.25 (from 7.2% to 6.95%) in T2D cases. This analysis will be conducted after the completion of ascertainment efforts for the 2017 incident cohort.

5.6.4 Aim 4: Adjusting Results Using Capture-Recapture Analysis

The completeness of ascertainment for each site will be estimated by dividing the number of identified cases by the estimated total number obtained from the capture-recapture analysis. The capture-recapture corrected estimate will be computed by dividing the observed incidence rate by the estimated capture-recapture rate. This corrected estimate can be seen as a ratio of 2 random variables. Pooled estimates that borrow information across site, sex and age groups will be used to guarantee that the capture-recapture rate and its associated standard error can be computed for all combinations of the variables considered in the analysis. Stratification by site, diabetes type, race/ethnicity, sex and age group can sometimes lead to small cell count causing convergence failures in the maximum likelihood estimation routines. Pooled estimation performed assuming a log-linear model makes it possible to obtain the maximum likelihood estimates in these cases and simplifies the derivation of the standard error associated with the estimated percentage completeness.

5.7 COHORT STUDY - STATISTICAL CONSIDERATIONS

5.7.1 Aim 1: Burden of Complications

Three main analytic approaches will be employed to examine the prevalence, incidence, progression and clustering of complications by diabetes type and responsible risk factors and pathways: a) estimating incidence and prevalence using multiple logistic regression methods; b) estimating the progression of complications using longitudinal mixed models; and c) estimating the clustering of risk factors and outcomes using longitudinal mixed models. For each of these approaches we will incorporate participant level

characteristics, measured at multiple time points, to examine potential mediators and moderators of outcomes.

Incidence Rate Estimation: Participants have had at least two previous in-person visits (C_0, C_1); however, for many outcomes (retinopathy, neuropathy, etc.) participants will have had only one previous assessment (C_1 visit). For these outcomes we will be able to define a group of participants who were free from the event of interest (e.g. no retinopathy) at C_1. Multiple logistic regression methods will be employed to examine the incidence rates of binary measures, with categorical (e.g.T1D vs. T2D) or continuous (e.g. A1c) predictors. We will evaluate potential confounding and/or effect modification based on our extensive databases.

Prevalence Rate Estimation: Some of the outcomes of interest have not been measured previously (e.g. echocardiography); therefore, we will estimate the prevalence of these outcomes. Associations of risk factors and diabetes type with prevalent outcomes will be examined using logistic regression models.

Statistical Power: For each of the primary dichotomous outcomes of interest (incidence or prevalence) we estimated proportions that will have the specific outcomes of interest, based on data from the C_1 visit. Table 6 shows the expected sample sizes available for comparing T1D and T2D, the corresponding detectable differences in rates, and the power for each outcome. We also provide the expected detectable differences in prevalence rates of LV hypertrophy between T1D and T2D in the sample of patients receiving echocardiography. These calculations are performed using Fisher's exact tests with $\alpha = 0.05$ (2-sided).

Table 6. Power for detectable differences for primary outcomes

Outcome	T1D/T2D available	T1D rate	T2D rate	Power
Incidence Comparisons				
Retinopathy	1215/230	10%	18%	88%
Neuropathy	1344/283	5%	10%	81%
Nephropathy	1172/268	20%	29%	85%
Prevalence Comparisons (Cardiac echocardiography)				
LV hypertrophy	250/250	5%	13%	86%

Longitudinal Models: We will use a longitudinal mixed effects analysis of covariance approach to make comparisons among groups which includes duration of diabetes as a time-varying covariate and participant as a random effect. This approach models the

varying durations of disease prior to the initial visit, and the varying durations of time allowed by the data collection windows between visits. These mixed effects models are flexible to allow for non-linear relationships to be modeled over time, and permit random rates of progression.

To estimate the sample size needed to detect a significant difference with sufficient power, calculations were based on comparing measurements after adjusting for C_0 or C_1 data. If the correlation between measurements is moderate (0.50) then we have 80% power to detect a difference of 0.139 standard deviations (SD) for each outcome of interest. For example, based on data collected on a subset of SEARCH T1D participants, the standard deviation for PWV carotid-femoral was estimated to be 0.7 m/s, thus we would have 80% power to detect a difference of 0.10 m/s in progression of PWV between youth with T2D vs T1D.

Clustering of Outcomes: In addition to examining each endpoint separately, we have the opportunity to look simultaneously at several outcomes in the same analysis. We will create variables that describe the co-occurrence (clustering) of outcomes for each participant and examine whether there are differences in the patterns of these clusters between T1D and T2D youth. Approaches will utilize ordinal logistic regression methods or longitudinal mixed models depending on whether the clustering outcome is a count or categorical. More sophisticated statistical methods may also be used such as principal components analyses to determine which risk factors may form different components.

5.7.2 Aim 2: Processes of Care

Analytically, the approach for addressing the questions related to processes of care, their influence on quality of life (QOL) during transition from pediatric to adult care, by diabetes type, will follow closely the approach described above for longitudinal models. For some analyses we need to assess the potential effects of mediators on the examined relationships. Potential mediator data has been measured in at least 3 time points. Furthermore, since the Affordable Care Act was implemented during the time frame when data has been collected, we can examine changes in outcomes that occur before or after that period.

For power calculations we will conservatively estimate a total of 2000 T1D and 382 T2D participants. If we assume that the correlation from the initial assessment of the outcome and the final assessment of the outcome is 0.5 then we can detect an effect size of 0.135 SD with 80% power (alpha=0.05, 2-sided). Thus, for an instrument such as the QOL scale where estimates of the standard deviation range from 13 to 17 this would correspond to having sufficient power to detect a difference between groups of 1.8 to 2.3 units, which is a clinically meaningful difference.

5.7.3 Aim 3: Mortality

We will perform both direct and indirect standardization to compare the death rates observed in SEARCH to the age-, race- and sex- matched US population and calculate the Standardized Mortality Ratio (SMR) for each subgroup. Statistical inference will be based on confidence interval estimation and Wald tests. SEARCH data will be used to compare mortality by diabetes type. Time to all-cause mortality will serve as the primary outcome for these analyses, and will be modeled using Cox proportional hazards.

The death rate estimated in the SEARCH 2002-2008 incident cohorts was 91.3 per 100,000 person-years. Based on projections, we expect to observe ~130 deaths in 142,000 person-years by 12/31/2017 (~82 among T1D and ~48 among T2D cases). Assuming alpha=0.05, we will have over 80% power to detect a hazard ratio (HR) of 1.9 or higher. With an observed HR for time to all-cause death of 2.7 in T2D relative to T1D currently, the proposed study will be well-powered to identify differential effect of diabetes type on mortality.

7. Study Organization

The organizational structure of SEARCH 4 is patterned after the successful structure of the previous phases of the study. The Steering Committee is the main governing body, and includes the Principal Investigators from each study site, the central laboratory, and the Coordinating Center; the chair of the Project Managers Committee; and the Project Scientists from the Center for Disease Control (CDC)/ National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP) and the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)**.** When voting is required, each site has one vote (five total), the CC has one vote, the funding agencies have one vote (combined), and the Project Manager Chair has one vote. Two co-chairs are selected from the non-federal Steering Committee members. The Steering Committee has primary responsibility to develop a common research protocol and manual of operations, facilitating the conduct and monitoring of the study, and reporting study results. The Steering Committee also oversees policies for access to participant data and specimens and ancillary studies. A Study Group is comprised of all Steering Committee members, plus additional investigators as well as consultants and project managers from the clinical sites and the CC. Key operational committees report directly to the Steering Committee.

An Observational Study Monitoring Board (OSMB) consisting of appropriately qualified independent experts provides review of data on study progress and participant safety. The purpose of the board is to assure independent review as to whether study participants are exposed to unreasonable risk because of study participation, and to monitor study progress and integrity. Board members are chosen by NIDDK, and typically convene twice a year (every 6 months) unless a need arises. The CC produces a report for review by the OSMB according to pre-determined format, contents, and reporting frequency. The reports present information regarding

(1) adverse events and safety violations experienced by study patients as a result of undergoing the study procedures and (2) conduct of the study, including withdrawals and visit attendance.

8. Quality Control

The SEARCH Coordinating Center is responsible for developing and implementing quality control (QC) procedures. QC techniques are incorporated into each phase of the study from case ascertainment, recruitment and registration of persons with diabetes through data acquisition, reading and/or interpretation of the results and their analyses and publication. The Coordinating Center continues to work with the QC Committee reporting to the Steering Committee and to the OSMB. The QC Committee works in concert with the Coordinating Center to oversee the standardized measurement protocols for collecting data during clinic visits and interviews. The committee oversees and recommends any revisions to, or further development of, study data collection forms; develops guidelines for and oversees the central laboratory and reading centers; reviews and monitors quality control related to study measures; and reports on quality control to the study group. This committee also reviews the certification of clinic staff and assists with training and certification/ recertification of study staff on measurement protocols. Any problems identified with laboratory and reading centers or clinic performance are addressed with remediation plans.

9. Centralized Data Management System

The SEARCH study features an integrated web-based system for managing operations and capturing data as developed by the CC. Once entered, data are immediately validated against sets of rules. Some of these rules identify errors that must be corrected immediately; others present validation warnings for review which are saved to the database for later reconciliation. Data are immediately available in alert/tracking systems and dynamic reports based on relational databases. No records are ever deleted, all changes produce audit trails, and back-ups are created hourly. This provides a high degree of integrity, detail, and flexibility in responding to unexpected study needs related to report generation, auditing, and monitoring. A comprehensive security program is in place that integrates policy and practice (see Appendix A).

The system allows authorized users to access clinic and participant information for the purpose of entering and editing study data. Only authorized users may access and enter/update information regarding participants' study data. Only local site staff and investigators and authorized Coordinating Center staff have access to data from individual sites. A correct username and password is required to gain access to the system and role-based security is employed to restrict user access to only authorized areas and data. All data are stored in a secured Microsoft SQL Server (2008) database system at Wake Forest School of Medicine. The system employs audit logs that capture and store each version of every record that is saved on the system. Users who access the system, once authenticated, establish a secure SSL encrypted session and all transmissions are encrypted until they logout or close the browser. The system is backed up nightly onto dedicated backup storage equipment.

10. Confidentiality

As in previous phases of SEARCH, every precaution is taken to maintain the confidentiality of all study participants. For both the Cohort and Registry Studies, confidentiality of data is maintained by using research identification (ID) numbers that uniquely identify each individual. Hardcopies of individual participants' research records will be retained and secured by each SEARCH Clinical Center. The file that links participants' names and demographic information with their research ID numbers is retained separately from the study data, using an approach consistent with local IRB requirements. After the study is completed, local data are stored with that of other completed studies in a secure storage area following all applicable local regulations for the storage, maintenance, and destruction of research data.

As in previous phases of the SEARCH Study, an NIH Certificate of Confidentiality is maintained at the CC to offer further protection of privacy.

11. Safety Management

The potential risks to individuals participating in the SEARCH 4 Cohort and Registry study components are very few. Participant safety is monitored through center specific guidelines. Study-related adverse events are documented on the Event Reporting Form and submitted to the Coordinating Center. An external reviewer reviews all events reported in this manner and reports findings to the SEARCH Quality Control Committee. The risks are described below along with strategies that are used to minimize these risks.

Blood samples

To minimize the possibility of risks associated with phlebotomy experienced medical staff obtain the blood samples in accordance with local guidelines. A numbing medicine may be placed on the skin before the blood is drawn to decrease any pain. Participants who have a history of fainting or who develop symptoms of light-headedness may be placed in the supine position and blood sugar levels are checked with a blood glucose meter.

Results reporting

Participants (or their parent/guardian if <18 years of age) are given all clinically relevant test results based on measurements and samples collected during their study visits. Transmission of results is based on the age of the participant at the time that the results become available. If the participant's parent agreed to have the samples drawn but the participant is at least 18 years of age when the results become available, then the participant is notified of the results.

Participants (or their parent/guardian if <18 years) are asked whether or not they wish their diabetes and/or primary care provider(s) to receive their clinically relevant test results such as HbA1c, glucose, lipid profile, C-peptide, diabetes autoantibodies, and urine albumin and creatinine. Receipt of these results is viewed as a possible but not definite benefit to the participant as such information may or may not affect subsequent diabetes (or complication) management. If critical laboratory values do occur, the central laboratory contacts the local Principal Investigator

and/or his/her designee, and the information is shared with the participant or his/her parent/guardian if <18 years of age, as well as the provider if permission was given at the time of the study visit. Participants with abnormalities needing medical management are referred to their primary care provider (PCP).

Information from interviews is not to be shared with parents or guardians with the exception of the Centers for the Epidemiologic Studies of Depression (CES-D) scale results that are at or above the alert value.

Identification of Alert Values

The following components of the Registry Study IPV and the Cohort Study exam have identified alert levels and a detailed action plan in the Manual of Procedures:

- serum glucose level < 45 mg/dl or > 300 mg/dl;
- triglyceride levels >1000 mg/dl;
- blood pressure > the 95th percentile;
- urine albumin:creatinine ratio ≥ 30µg/mg
- untreated ulcer or infection of feet;
- pathology identified on retinal photography;
- pathology identified on cardiac echocardiography;
- elevated CES-D total score: > 24 for participants < 18 yrs. of age and ≥ 16 for participants ≥ 18 yrs.

APPENDIX B

Information System Security Plan for Wake Forest/Public Health Sciences

General System Description of Data Management System

The SEARCH data management system allows only authorized users to access participant information and enter/update information regarding participants' study data. The application maintains audit logs which identify the activity of each user at all times while logged into the system. This system is built as a web-based application which is accessed via the Internet. A correct username and password is required to gain access to the system and role based security is employed to restrict user access to only authorized areas and data. The application is built using HTML forms and Macromedia's ColdFusion middleware product for database interactions. Javascript and a ColdFusion based rules engine provides data validation and integrity checking on all submitted data. All data is stored in a secured Microsoft SQL Server (2008) database system. The system employs audit logs that capture and store each version of every record that is saved on the system. Users who access the system, once authenticated, establish a secure SSL encrypted session and all transmissions are encrypted until they logout or close the browser.

System Environment

The system is comprised of a Microsoft-based web server which runs Adobe's ColdFusion application server for integration of the database information with the web site. All data resides in a Microsoft SQL Server database with the appropriate role-based security maintained on the data. The application itself also implements role-based security to prevent unauthorized access to or manipulation of confidential information. The system is backed up nightly onto dedicated backup storage equipment. The application is hosted on a virtual server using VMWare. The server is in a secure DMZ zone. The server is maintained as all other servers in a secure data center and updated monthly with patches to the operating system and to the VMWare software. The server is backed up nightly and is on a UPS in the event of a power failure.

Backups

Nightly backups, moved offsite regularly, are made of all data and stored in secure fireproof cabinets. The backup schedule consists of full monthly backups and nightly incremental backups. Backup tapes are handled by two system administrators. Tapes are transported by one of two identified tape custodians. The tapes are moved from the data center to the offsite storage facility and are stored in fireproof cabinets. At all times during the transport, one of the tape custodians is present with the tapes. Tapes are identified by unique bar code labels accessible only by the systems administrators. This is the only information on the tape label. The backup system stores the information for each bar code with details of directories/files backed up that includes the date and time of backup. The backup system, when needing to restore files, will identify which tape is needed based on the bar code label. Only designated system administrators can restore the backup tapes.

Server Management and Data Center

The servers involved in this project are contained within a secure Data Center with environmental controls which detect abnormal conditions such as power outages, high heat or humidity, and loud sound. In the event of an abnormal condition, the system contacts three (3) individuals to notify them of the alerts. The Data Center has several secure access points that are accessible only by a badge reader. Only authorized staff will have accessible badges to these areas. The building is surrounded by a 10 foot fence with a gate access through badge control. The outside building door is accessed through badge control. The data center room is housed in a locked computer room that is accessed through badge control. Each

of these access controls is in place 24 hours a day and seven days a week. All servers have uninterruptible power supplies (UPS). The building has a backup generator that will automatically initiate in the event of a power failure. The computer room is equipped with fire suppression equipment. This equipment is tested on a scheduled timetable by the institution. The entire Data Center is fire-protected by a clean agent system which is backed up by a dry-pipe pre-action sprinkler system. The Data Center room is located on the second floor of the building in an area with no windows and has a raised floor to protect against flooding. The system is protected by a Cisco firewall and is located in a secure DMZ. Servers are protected by institution supported and maintained intrusion detection software as well as by SecureIIS which monitors incoming server traffic.

Password Security

Minimum password requirements must meet Wake Forest Health Sciences Security Policy requirements of:

- Must be changed every 90 days
- Administrative level passwords must change every 30 days
- Must be at least six characters long
- Must include any three of the following items
- English uppercase characters (A through Z)
- English lowercase characters (a through z)
- Numerals (0 - 9)
- Special characters (!, $, #, %, @, etc.)
- The same password cannot be reused in less than 4 previous passwords.

Code Scanning/Testing

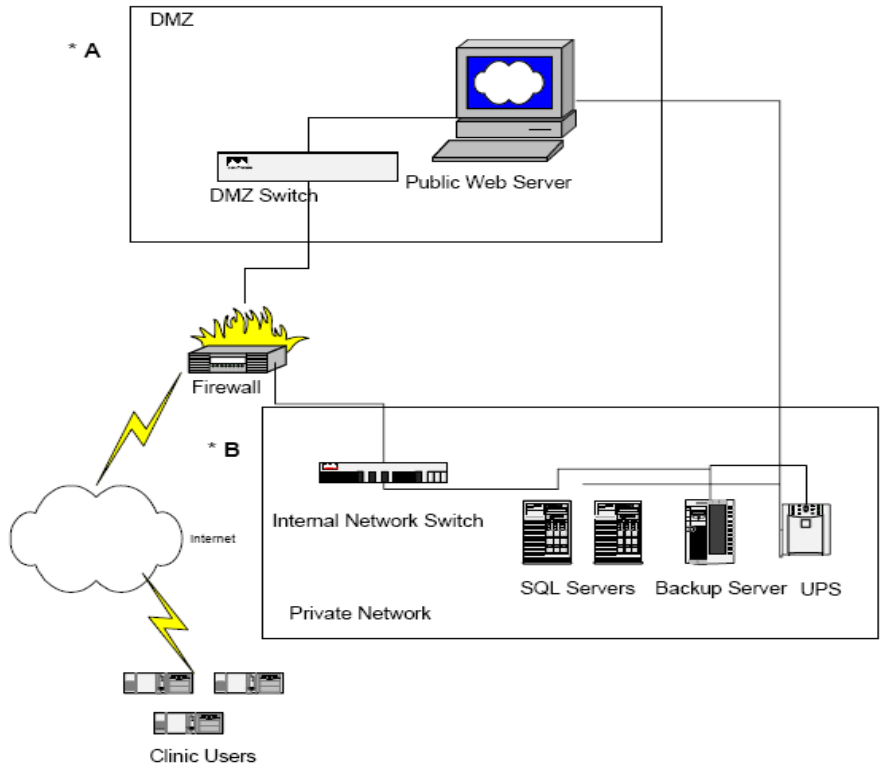
Prior to the release of the web site for public access, the Security Office scans the site for vulnerabilities such as, but not limited to, cross-site scripting, SQL injection attacks, and unsecured logins. The vulnerabilities are classified into five categories of Critical, High, Medium, Low and Best Practices. All Critical and High vulnerabilities must be resolved. Each medium and low vulnerability is reviewed and after discussion with the Security Office, decisions are made to remediate the issue or that the issue is not a security risk to the organization. The Security Office uses the WebInspect product from HP. The tool is automatically updated at each scan for new vulnerabilities. The web site is scanned at the initial release and at least annually thereafter. If significant changes have been made to the site, the site is required to undergo additional scans prior to the annual scan.

Disaster and Contingency Planning

Hurricanes Katrina (ACCORD) and Sandy (SPRINT) have made clear the need for careful disaster planning. While our CCs were not directly impacted by these acts of nature, each forced a clinical site to close (at least temporarily). The Department of Biostatistical Sciences has a disaster plan as part of our NHLBI-approved information security plan. This plan identifies key personnel that need to be notified in times of disaster as well as which critical systems need to be brought online first. The plan describes how we would continue business operations should a disaster happen by identifying alternative human and computational resources that we could leverage should a disaster strike.

System/Network Diagram

High Level Information System Diagram



Reference List

1. Primary Care of the Child with a Chronic Condition. 4 ed. St. Louis: Mosby-Yearbook; 2004.

2. Pambianco G, Costacou T, Ellis D, Becker DJ, Klein R, Orchard TJ. The 30-year natural history of type 1 diabetes complications: the Pittsburgh Epidemiology of Diabetes Complications Study experience. *Diabetes* 2006;55(5):1463-1469.

3. Patterson C, Guariguata L, Dahlquist G, Soltesz G, Ogle G, Silink M. Diabetes in the young - a global view and worldwide estimates of numbers of children with type 1 diabetes. *Diabetes Res Clin Pract* 2014;103(2):161-175.

4. Patterson CC, Gyurus E, Rosenbauer J, Cinek O, Neu A, Schober E, Parslow RC, Joner G, Svensson J, Castell C, Bingley PJ, Schoenle E, Jarosz-Chobot P, Urbonaite B, Rothe U, Krzisnik C, Ionescu- Tirgoviste C, Weets I, Kocova M, Stipancic G, Samardzic M, de Beaufort CE, Green A, Dahlquist GG, Soltesz G. Trends in childhood type 1 diabetes incidence in Europe during 1989-2008: evidence of non- uniformity over time in rates of increase. *Diabetologia* 2012;55(8):2141-2147.

5. The Writing Group for the SEARCH for Diabetes in Youth Study Group. Incidence of Diabetes in Youth in the United States. *JAMA: The Journal of the American Medical Association* 2007;297(24):2716-2724.

6. Pettitt DJ, Talton J, Dabelea D, Divers J, Imperatore G, Lawrence JM, Liese AD, Linder B, Mayer-Davis EJ, Pihoker C, Saydah SH, Standiford DA, Hamman RF, for the SEARCH for Diabetes in Youth Study Group. Prevalence of Diabetes Mellitus in U.S. Youth in 2009: The SEARCH for Diabetes in Youth Study. *Diab Care* 2014;37(1):402-408.

7. Patterson CC, Gyurus E, Rosenbauer J et al. Trends in childhood type 1 diabetes incidence in Europe during 1989-2008: evidence of non-uniformity over time in rates of increase. *Diabetologia* 2012 August;55(8):2142-7.

8. Lawrence JM, Imperatore G, Dabelea D et al. Trends in incidence of type 1 diabetes among non-Hispanic white youth in the U.S., 2002-2009. *Diabetes* 2014 November;63(11):3938-45.

9. Dabelea D, Mayer-Davis EJ, Saydah S et al. Prevalence of type 1 and type 2 diabetes among children and adolescents from 2001 to 2009. *JAMA* 2014 May 7;311(17):1778-86.

10. Harjutsalo V, Sund R, Knip M, Groop PH. Incidence of type 1 diabetes in Finland. *JAMA* 2013 July 24;310(4):427-8.

11. Pinhas-Hamiel O, Zeitler P. "Who is the wise man?--The one who foresees consequences:". Childhood obesity, new associated comorbidity and prevention. *Prev Med* 2000 December;31(6):702-5.

12. SEARCH Study Group. SEARCH for Diabetes in Youth: a multicenter study of the prevalence, incidence and classification of diabetes mellitus in youth. *Control Clin Trials* 2004 October;25(5):458-71.

13. Hamman RF, Bell RA, Dabelea D et al. The SEARCH for Diabetes in Youth study: rationale, findings, and future directions. *Diabetes Care* 2014 December;37(12):3336-44.

14 Mayer-Davis EJ, Bell RA, Dabelea D, D'Agostino R, Jr., Imperatore G, Lawrence JM, Liu L, Marcovina S, SEARCH for Diabetes in Youth Study Group. The many faces of diabetes in American youth: type 1 and type 2 diabetes in five race and ethnic populations: the SEARCH for Diabetes in Youth Study. Diab Care 2009;32 Suppl 2:S99-101.

15. Dabelea D, Pihoker C, Talton JW, D'Agostino RB, Jr., Fujimoto W, Klingensmith GJ, Lawrence JM, Linder B, Marcovina SM, Mayer-Davis EJ, Imperatore G, Dolan LM. Etiological Approach to Characterization of Diabetes Type: The SEARCH for Diabetes in Youth Study. *Diab Care* 2011;34(7):1628-1633.

16. Hamman, R.F. et al (2013). Estimation of completeness of case ascertainment using capture-recapture. Unpublished report. SEARCH For Diabetes in Youth: a multi-center study. https://www.searchfordiabetes.org/public/dsphome.cfm

17. Lawrence JM; Black MH; Zhang JL; Slezak JM; Takhar HS; Koebnick C, Mayer-Davis EJ; Zhong VW; Dabelea D; Hamman RF; Reynolds K. Validation of Pediatric Diabetes Case Identification Approaches for Diagnosed Cases Using Information in the Electronic Health Records of a Large Integrated Managed Health Care Organization. Am J Epidemiol. 2014 Jan 1;179(1):27-38. doi: 10.1093/aje/kwt230. Epub 2013 Oct 7.

18. Doody MM, Hayes HM, Bilgrad R. Comparability of National Death Index Plus and Standard Procedures for Determining Causes of Death in Epidemiologic Studies. *Ann Epidemiol* 2001;11(1):46- 50.

19. Tamborlane WV, Beck RW, Bode BW, Buckingham B, Chase HP, Clemons R, Fiallo-Scharer R, Fox LA, Gilliam LK, Hirsch IB, Huang ES, Kollman C, Kowalski AJ, Laffel L, Lawrence JM, Lee J, Mauras N, O'Grady M, Ruedy KJ, Tansey M, Tsalikian E, Weinzimer S, Wilson DM, Wolpert H, Wysocki T, Xing D. Continuous glucose monitoring and intensive treatment of type 1 diabetes. *N Engl J Med* 2008;359(14):1464-1476.

20. Beck RW, Buckingham B, Miller K, Wolpert H, Xing D, Block JM, Chase HP, Hirsch I, Kollman C, Laffel L, Lawrence JM, Milaszewski K, Ruedy KJ, Tamborlane WV. Factors predictive of use and of benefit from continuous glucose monitoring in type 1 diabetes. *Diab Care* 2009;32(11):1947-1953.

21. Mayer-Davis EJ, Nichols M, Liese AD, Bell RA, Dabelea DM, Johansen JM, Pihoker C, Rodriguez BL, Thomas J, Williams D. Dietary intake among youth with diabetes: the SEARCH for Diabetes in Youth Study. *J Am Diet Assoc* 2006;106(5):689-697.

22. O'Neill JR, Liese AD, McKeown RE, Cai B, Cuffe SP, Mayer-Davis EJ, Hamman RF, Dabelea D. Physical Activity and Self-Concept: The SEARCH for Diabetes in Youth Case Control Study. *Pediatr Exerc Sci* 2012;24(4):577-588.

23. Li C, Beech B, Crume T, D'Agostino RB, Jr., Dabelea D, Kaar JL, Liese AD, Mayer-Davis EJ, Pate R, Pettitt DJ, Taplin C, Rodriguez B, Merchant AT. Longitudinal association between television watching and computer use and risk markers in diabetes in the SEARCH for Diabetes in Youth Study. *Pediatr Diabetes* 2014;10.1111/pedi.12163.

24. Reynolds K, Liese AD, Anderson AM, Dabelea D, Standiford D, Daniels SR, Waitzfelder B, Case D, Loots B, Imperatore G, Lawrence JM. Prevalence of tobacco use and association between cardiometabolic risk factors and cigarette smoking in youth with type 1 or type 2 diabetes mellitus. *J Pediatrics* 2011;158(4):594-601.

25. Winters KC. Development of an adolescent alcohol and other drug abuse screening scale: Personal experience screening questionnaire. *Addictive Behaviors* 1992;17(5):479-490.

26. Lawrence JM, Standiford DA, Loots B, Klingensmith GJ, Williams DE, Ruggiero A, Liese AD, Bell RA, Waitzfelder BE, McKeown RE. Prevalence and correlates of depressed mood among youth with diabetes: the SEARCH for Diabetes in Youth study. *Pediatrics* 2006;117(4):1348-1358.

27. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *App Psych Measure* 1977;1:385-401.

28. Hood KK, Lawrence JM, Anderson A, Bell R, Dabelea D, Daniels S, Rodriguez B, Dolan LM. Metabolic and inflammatory links to depression in youth with diabetes. *Diab Care* 2012;35(12):2443-2446.

29. Olino TM, Yu L, McMakin DL, Forbes EE, Seeley JR, Lewinsohn PM, Pilkonis PA. Comparisons across depression assessment instruments in adolescence and young adulthood: an item response theory study using two linking methods. *J Abnorm Child Psychol* 2013;41(8):1267-1277.

30. Cox DJ, Irvine A, Gonder-Frederick L, Nowacek G, Butterfield J. Fear of hypoglycemia: quantification, validation, and utilization. *Diab Care* 1987;10(5):617-621.

31. Green LB, Wysocki T, Reineck BM. Fear of hypoglycemia in children and adolescents with diabetes. J *Pediatr Psychol* 1990;15(5):633-641.

32. Clarke WL, Gonder-Frederick A, Snyder AL, Cox DJ. Maternal fear of hypoglycemia in their children with insulin dependent diabetes mellitus. *J Pediatr Endocrinol Metab* 1998;11 Suppl 1:189-194.

33. Gonder-Frederick L, Nyer M, Shepard JA, Vajda K, Clarke W. Assessing fear of hypoglycemia in children with Type 1 diabetes and their parents. *Diabetes Manag* (Lond) 2011;1(6):627-639.

34. Hood KK, Butler DA, Anderson BJ, Laffel LMB. Updated and Revised Diabetes Family Conflict Scale. *Diab Care* 2007;30(7):1764-1769.

35. Schmitt A, Gahr A, Hermanns N, Kulzer B, Huber J, Haak T. The Diabetes Self-Management Questionnaire (DSMQ): development and evaluation of an instrument to assess diabetes self-care activities associated with glycemic control. *Health Qual Life Outcomes* 2013;11:138.

36. Agency for Health Care Research and Quality. CAHPS Item Set for Children with Chronic Conditions. https://cahps ahrq gov/surveys-guidance/item-sets/children-chronic/index html, 2014 (Accessed December 17, 2014 https://cahps.ahrq.gov/surveys-guidance/item-sets/children-chronic/index.html).

37. American Diabetes Association. Diagnosis and Classification of Diabetes Mellitus. *Diab Care* 2014;37(Supplement 1):S81-S90.

38. Naughton MJ, Ruggiero AM, Lawrence JM, Imperatore G, Klingensmith GJ, Waitzfelder B, McKeown RE, Standiford DA, Liese AD, Loots B, SEARCH for Diabetes in Youth Study Group. Health-related quality of life of children and adolescents with type 1 or type 2 diabetes mellitus: SEARCH for Diabetes in Youth Study. *Arch Pediatr Adolesc Med* 2008;162(7):649-657.

39. Lawrence JM, Yi-Frazier JP, Black MH, Anderson A, Hood K, Imperatore G, Klingensmith GJ, Naughton M, Mayer-Davis EJ, Seid M. Demographic and Clinical Correlates of Diabetes-Related Quality of Life among Youth with Type 1 Diabetes. *J Pediatrics* 2012;161(2):201-207.

40. Naughton MJ, Yi-Frazier JP, Morgan TM, Seid M, Lawrence JM, Klingensmith GJ, Waitzfelder B, Standiford DA, Loots B. Longitudinal associations between sex, diabetes self-care, and health-related quality of life among youth with type 1 or type 2 diabetes mellitus. *J Pediatrics* 2014;164(6):1376-1383.

41. Ngo-Metzger Q, Sorkin DH, Billimek J, Greenfield S, Kaplan SH. The effects of financial pressures on adherence and glucose control among racial/ethnically diverse patients with diabetes. *J Gen Intern Med* 2012;27(4):432-437.

42. Coleman EA, Mahoney E, Parry C. Assessing the quality of preparation for posthospital care from the patient's perspective: the care transitions measure. *Med Care* 2005;43(3):246-255.

43. Lang RM, Bierig M, Devereux RB, Flachskampf FA, Foster E, Pellikka PA, Picard MH, Roman MJ, Seward J, Shanewise JS, Solomon SD, Spencer KT, Sutton MS, Stewart WJ. Recommendations for chamber quantification: a report from the American Society of Echocardiography's Guidelines and Standards Committee and the Chamber Quantification Writing Group, developed in conjunction with the European Association of Echocardiography, a branch of the European Society of Cardiology. *J Am Soc Echocardiogr* 2005;18(12):1440-1463.

44. Gottdiener JS, Bednarz J, Devereux R, Gardin J, Klein A, Manning WJ, Morehead A, Kitzman D, Oh J, Quinones M, Schiller NB, Stein JH, Weissman NJ. American Society of Echocardiography recommendations for use of echocardiography in clinical trials. *J Am Soc Echocardiogr* 2004;17(10):1086-1119.

45. Devereux RB, Alonso DR, Lutas EM, Gottlieb GJ, Campo E, Sachs I, Reichek N. Echocardiographic assessment of left ventricular hypertrophy: comparison to necropsy findings. *Am J Cardiol* 1986;57(6):450-458.

46. National Center for Health Statistics. National Health and Nutrition and Examination Survey (NHANES) Ophthalmology Procedures Manual. 2005 Sep; www.cdc.gov/nchs/data/nhanes/nhanes_05_06/OP.pdf<http://www.cdc.gov/nchs/data/nhanes/nhanes_05_06/OP.pdf>.

47. Klein BE, Knudtson MD, Tsai MY, Klein R. The relation of markers of inflammation and endothelial dysfunction to the prevalence and progression of diabetic retinopathy: Wisconsin epidemiologic study of diabetic retinopathy. *Arch Ophthalmol* 2009;127(9):1175-1182.

48. Schwartz GJ, Schneider MF, Maier PS, Moxey-Mims M, Dharnidharka VR, Warady BA, Furth SL, Munoz A. Improved equations estimating GFR in children with chronic kidney disease using an immunonephelometric determination of cystatin C. *Kidney Int* 2012;82(4):445-453.

49. Inker LA, Shaffi K, Levey AS. Estimating glomerular filtration rate using the chronic kidney disease- epidemiology collaboration creatinine equation: better risk predictions. *Circ Heart Fail* 2012;5(3):303- 306.

50. Sharma AP, Yasin A, Garg AX, Filler G. Diagnostic Accuracy of Cystatin C–Based eGFR Equations at Different GFR Levels in Children. *Clin J Am Soc Nephro* 2011; 6: 1599–1608.
51. Bouvet Y, Bouissou F, Coulais Y, Seronie-Vivien S, Tafani M, Decramer S, et al: GFR is better estimated by considering both serum cystatin C and creatinine levels. *Pediatr Nephrol* 2006; 21: 1299–1306

52. Feldman EL, Stevens MJ, Thomas PK, Brown MB, Canal N, Greene DA. A practical two-step quantitative clinical and electrophysiological assessment for the diagnosis and staging of diabetic neuropathy. *Diab Care* 1994;17(11):1281-1289.

53. Moghtaderi A, Bakhshipour A, Rashidi H. Validation of Michigan neuropathy screening instrument for diabetic peripheral neuropathy. *Clin Neurol Neurosurg* 2006;108(5):477-481.

54. Herman WH, Pop-Busui R, Braffett BH, Martin CL, Cleary PA, Albers JW, Feldman EL, The DCCT/EDIC Research Group. Use of the Michigan Neuropathy Screening Instrument as a measure of distal symmetrical peripheral neuropathy in Type-á1 diabetes: results from the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications. *Diabetic Med* 2012;29(7):937-944.

55. Wadwa RP, Urbina EM, Anderson AM, Hamman RF, Dolan LM, Rodriguez BL, Daniels SR, Dabelea D. Measures of arterial stiffness in youth with type 1 and type 2 diabetes: the SEARCH for Diabetes in Youth study. *Diab Care* 2010;33(4):881-886.

56. Lang RM, Bierig M, Devereux RB, Flachskampf FA, Foster E, Pellikka PA, Picard MH, Roman MJ, Seward J, Shanewise JS, Solomon SD, Spencer KT, Sutton MS, Stewart WJ. Recommendations for chamber quantification: a report from the American Society of Echocardiography's Guidelines and Standards Committee and the Chamber Quantification Writing Group, developed in conjunction with the European Association of Echocardiography, a branch of the European Society of Cardiology. *J Am Soc Echocardiogr* 2005;18(12):1440-1463.

57. NIH Toolbox for the assessment of neurological and behavioral function. http://www.nihtoolbox.org/Pages/default aspx, 2014 (Accessed December 8, 2015).

58. The DCCT research group. Epidemiology of severe hypoglycemia in the diabetes control and complications trial. *Am J Med* 1991;90(4):450-459.

59. Svoren BM, Volkening LK, Butler DA, Moreland EC, Anderson BJ, Laffel LM. Temporal trends in the treatment of pediatric type 1 diabetes and impact on acute outcomes. *J Pediatr* 2007;150(3):279-285.

60. Zhong VW, Pfaff ER, Beavers DP et al. Use of administrative and electronic health record data for development of automated algorithms for childhood diabetes case ascertainment and type classification: the SEARCH for Diabetes in Youth Study. *Pediatr Diabetes* 2014 December;15(8):573-84.

61. Dabelea D, DeGroat J, Sorrelman C et al. Diabetes in Navajo youth: prevalence, incidence, and clinical characteristics: the SEARCH for Diabetes in Youth Study. *Diabetes Care* 2009 March;32 Suppl 2:S141-S147.

62. Elding Larsson H, Vehik K, Bell R, Dabelea D, Dolan L, Pihoker C, Knip M, Veijola R, Lindblad B,Samuelsson U, Holl R, Haller MJ, on behalf of the TEDDY Study Group, SEARCH Study Group, Swediabkids Study Group, DPV Study Group, Finnish Diabetes Registry Study Group. Reduced Prevalence of Diabetic Ketoacidosis at Diagnosis of Type 1 Diabetes in Young Children Participating in Longitudinal Follow-Up. Diab Care 2011;34(11):2347-2352.

APPENDIX C SEARCH Food Security Supplemental Visits Protocol

SEARCH Food Security Cohort Study

Supplemental Visit Protocol

Version 2.0: May 4, 2018
Version 3.0: October 17, 2018
Version 4.0: April 29, 2019
Version 5: July 30, 2019
Version 6: October 30, 2020
Version 7: December 15, 2021

Project Period: April 1, 2018 – March 31, 2023

APPENDIX C Table of Contents

List of Tables and Figures

Tables

Figures

SEARCH Food Security (SFS) Cohort Study

Executive Summary

This protocol describes the goals and methods of the SEARCH Food Security Cohort Study which was funded by the application entitled "Impact of Disparities in Food Security on Glycemic Control and Health Care Utilization Among Youth and Young Adults with Diabetes" to the National Institutes of Health.

Background and Aims. The overarching goal of the management of type 1 (T1D) and type 2 (T2D) diabetes is optimal glycemic control, which is key to reducing the risk of morbidity and preventing premature mortality.[1, 2] Despite advances in medication therapy, 55.6% of youth and young adults (YYAs) with T1D and 46% of those with T2D do not achieve optimal glycemic control.[3-5] Minority YYAs are particularly disadvantaged: 65% of non-Hispanic black and 61% of Hispanic YYAs with T1D do not have optimal glycemic control compared to 29% of non-Hispanic whites, and similar inequities exist in T2D (41% and 49% vs. 19%, respectively).[3]

The 2016 Standards of Care for Diabetes for the first time explicitly recommend assessing household food insecurity (HFI) because of its central role in influencing the three pillars of optimal glycemic control: nutrition therapy, physical activity, and glucose monitoring/self-management.[1] HFI is defined as "limited or uncertain availability of nutritionally adequate and safe foods…"[6] HFI not only is a nutritional hardship but also exerts severe negative influences on mental and physical health.[7] In persons with diabetes, the consequences of HFI can include poor glycemic control,[8, 9] hypoglycemia, and a much higher frequency of hospitalization.[8-12] In non-diabetics, HFI is associated with 56% higher total health care costs independent of other social determinants of health.[13] ***Because HFI disproportionately affects racial/ethnic minorities,[14, 15] it may underlie disparities in diabetes outcomes that disadvantage minorities***.[16]

Cross-sectional studies, mostly in older adults with T2D, inform the scientific premise of this study on the impact of HFI on glycemic control and health care utilization in persons with diabetes. However, virtually nothing is known about the impact of HFI on YYAs with T1D or T2D and YYAs of minority race/ethnicity or how individuals' behaviors and children's own experiences of HFI mediate the impact of HFI. Studies focusing specifically on YYAs with T1D and T2D are necessary because experiencing HFI in youth has unique negative influences, making extrapolation from older adults with T2D insufficient.[7] Moreover, no studies have examined HFI as a mechanism for disparities in adverse outcomes among minority YYAs with diabetes.

Thus, we propose a longitudinal study, the ***SEARCH Food Security (SFS) cohort study, which will build on the rich data of the SEARCH for Diabetes in Youth study***, to address the following aims:

Aim 1. To initiate the multi-center SFS cohort study of YYAs with T1D and T2D by leveraging the ongoing SEARCH 4 study and adding two subsequent data collection time points at three sites.

Aim 2. To prospectively evaluate the influence of HFI on changes in glycemic control in YYAs with T1D and T2D. We hypothesize that (1) HFI is associated with increases in HbA_{1c} values over ~18-27 months and more frequent episodes of hypoglycemia and diabetic ketoacidosis, independent of race/ethnicity and other social determinants of health; (2) these associations are stronger in racial/ethnic minority versus non-Hispanic white YYAs; (3) these association are stronger in T1D than T2D; and (4) these associations are buffered (moderated) by food assistance.

Aim 3. To quantify the mediating role of nutritional, mental health, and behavioral pathways through which HFI may affect changes in glycemic control in YYAs with T1D and T2D. Sub-aim 3.1. To evaluate self-reported food insecurity experiences of youth with T1D and understand their contribution to the mediating pathways, independent of HFI.

Aim 4. To prospectively evaluate the influence of HFI on changes in health care utilization and medical and non-medical health care costs in YYAs with T1D and T2D. We hypothesize that (1) HFI is associated with increases in health care utilization and costs, independent of other social determinants of health; (2) these associations are stronger in racial/ethnic minority versus non-Hispanic white YYAs; (3) these associations are stronger in YYAs with T1D than those with T2D; and (4) these associations are moderated by food assistance.

Methods.

The proposed study will leverage the ongoing SEARCH for Diabetes in Youth study (SEARCH 4, 2015-2020) to conduct a food insecurity–focused longitudinal ancillary study, the ***SFS cohort study***, that prospectively examines the association of food insecurity with glycemic control, health care utilization, and costs in YYAs with T1D and T2D. We will integrate SEARCH data with new SFS data. The SFS design was guided by our conceptual framework and developed to (1) comprehensively assess food insecurity in YYAs with diabetes; (2) optimize temporal clarity and separation between assessment of exposure (e.g., food insecurity), mediators (e.g., nutritional, mental health, diabetes self-management behaviors), and outcomes (e.g., glycemic control, health care utilization, medical and non-medical health care costs); (3) efficiently leverage the racial/ethnic diversity of SEARCH; and (4) include YYAs with T1D and T2D. The proposed work complements but does not duplicate SEARCH 4 or any SEARCH ancillary study and is supported by SEARCH.

The SFS cohort study will utilize data from three time points, including SEARCH 4 plus two SFS data collections spaced nine months after SEARCH 4. To maintain and maximize efficiency while maintaining excellent racial/ethnic diversity and numbers of participants with T2D, the SFS study will be conducted at three of the five SEARCH sites: Colorado (CO), South Carolina (SC), and Washington (WA).

The starting population for SFS is anticipated to be n=1,630 (999 white, 631 minority), anticipating 75% retention for follow-up 1 (FU 1 n=1,222; 654 white, 568 minority) and 90% retention for follow-up 2 (FU 2 n=962; 450 white, 512 minority). Data to be collected under the SFS cohort study will include (1) repeat assessment of food security and food assistance at FU 1 and 2, (2) information on hypothesized mediators (e.g., nutrition, mental health, diabetes self-management behaviors), (3) information on covariates (confounders) characterizing other social determinants of health not assessed in SEARCH 4 (i.e., literacy, homelessness), and (4) outcomes (e.g., HbA_{1c}, hypoglycemia, ketoacidosis, health care utilization and costs).

Because SEARCH 4 data collection started in July 2016, ~600 of the estimated 1,187 SFS-eligible persons will have 12-27 months of follow-up time between SEARCH 4 and SFS FU 1 (October 2018), the rest will have 9 months follow-up. SFS FU1 recruitment will be front-loaded, balancing rapid recruitment of those with longer follow-up times with maintaining the 9 months follow-up for the rest. FU 1 and 2 will be nine months apart. Thus, SFS FU 2 will be ~18-36 months after a participant's SEARCH 4 visit and/or survey completion. *Due to the Coronavirus, the FU2 clinic visit component was suspended in March 2020, to protect the health and safety of research study personnel and research participants. Because of the ongoing coronavirus pandemic, HbA1c will be collected from participants via dried blood spot instead of a blood draw in a clinic setting. The dried blood spot method can be remotely self-administered by the participant and returned by mail. Participants who prefer to have the dried blood spot administered in a clinic setting will be offered that option, if available. * SEARCH classification of diabetes type at recruitment uses provider-based information supplemented by diabetes autoantibodies (DAAs) and insulin sensitivity.[17]

To the best of our knowledge, the proposed study will be the first to characterize the unique HFI challenges faced and consequences experienced by YYAs with T1D and T2D in the US. HFI was recently recognized by the Centers for Medicare and Medicaid Services as highly clinically relevant.[18] Leveraging the existing SEARCH study will guarantee high scientific rigor, timeliness, and cost-efficiency. Our findings have the potential to alter policy and clinical practice by establishing the need for (1) more in-depth assessments of HFI among YYAs with diabetes than current screening provides, (2) reframing of diabetes management and nutrition recommendations for food-insecure patients, and (3) integration of food assistance resources into routine diabetes care for food-insecure YYAs, with potential emphasis on YYAs belonging to racial/ethnic minorities. ***In light of the high prevalence of household HFI in the US, which affected 12.7% of the population in 2015 (16.6% of households with children),[14] the proposed study is important and timely.***

Study Aims and Hypotheses

The overarching goal of the management of type 1 (T1D) and type 2 (T2D) diabetes is optimal glycemic control, which is key to reducing the risk of morbidity and preventing premature mortality.[1, 2] Despite advances in medication therapy, 55.6% of youth and young adults (YYAs) with T1D and 46% of those with T2D do not achieve optimal glycemic control.[3-5] Minority YYAs are particularly disadvantaged: 65% of non-Hispanic black and 61% of Hispanic YYAs with T1D do not have optimal glycemic control compared to 29% of non-Hispanic whites, and similar inequities exist in T2D (41% and 49% vs. 19%, respectively).[3]

The 2016 Standards of Care for Diabetes for the first time explicitly recommend assessing household food insecurity (HFI) because of its central role in influencing the three pillars of optimal glycemic control: nutrition therapy, physical activity, and glucose monitoring/self-management.[1] HFI is defined as "limited or uncertain availability of nutritionally adequate and safe foods…"[6] HFI not only is a nutritional hardship but also exerts severe negative influences on mental and physical health.[7] In persons with diabetes, the consequences of HFI can include poor glycemic control,[8, 9] hypoglycemia, and a much higher frequency of hospitalization.[8-12] In non-diabetics, HFI is associated with 56% higher total health care costs independent of other social determinants of health.[13] ***Because HFI disproportionately affects racial/ethnic minorities,[14, 15] it may underlie disparities in diabetes outcomes that disadvantage minorities.***[16]

Cross-sectional studies, mostly in older adults with T2D, inform the scientific premise of this study on the impact of HFI on glycemic control and health care utilization in persons with diabetes. However, virtually nothing is known about the impact of HFI on YYAs with T1D or T2D and YYAs of minority race/ethnicity or how individuals' behaviors and children's own experiences of HFI mediate the impact of HFI. Studies focusing specifically on YYAs with T1D and T2D are necessary because experiencing HFI in youth has unique negative influences, making extrapolation from older adults with T2D insufficient.[7] Moreover, no studies have examined HFI as a mechanism for disparities in adverse outcomes among minority YYAs with diabetes.

Thus, we propose a longitudinal study, the ***SEARCH Food Security (SFS) cohort study, which will build on the rich data of the SEARCH for Diabetes in Youth study***, to address the following aims:

Aim 1. To initiate the multi-center SFS cohort study of YYAs with T1D and T2D by leveraging the ongoing SEARCH 4 study and adding two subsequent data collection time points at three sites.

Aim 2. To prospectively evaluate the influence of HFI on changes in glycemic control in YYAs with T1D and T2D.

We hypothesize that:

(1) HFI is associated with increases in HbA_{1c} values over ~18-27 months and more frequent episodes of hypoglycemia and diabetic ketoacidosis, independent of race/ethnicity and other social determinants of health;

(2) these associations are stronger in racial/ethnic minority versus non-Hispanic white YYAs;

(3) these association are stronger in T1D than T2D; and

(4) these associations are buffered (moderated) by food assistance.

Aim 3. To quantify the mediating role of nutritional, mental health, and behavioral pathways through which HFI may affect changes in glycemic control in YYAs with T1D and T2D.

Sub-aim 3.1. To evaluate self-reported food insecurity experiences of youth with T1D and understand their contribution to the mediating pathways, independent of HFI.

Aim 4. To prospectively evaluate the influence of HFI on changes in health care utilization and medical and non-medical health care costs in YYAs with T1D and T2D.

We hypothesize that

(1) HFI is associated with increases in health care utilization and costs, independent of other social determinants of health;

(2) these associations are stronger in racial/ethnic minority versus non-Hispanic white YYAs;

(3) these associations are stronger in YYAs with T1D than those with T2D; and

(4) these associations are moderated by food assistance.

Background and Significance

Household food insecurity and standards of care: a new era

In 2016, the American Diabetes Association (ADA) and American Academy of Pediatrics (AAP) integrated food insecurity into their practice guidelines, recommending routine food insecurity screening in all persons with diabetes and all children and adolescents and these have been retained and expanded in subsequent iterations of the practice guidelines.[1, 19] They also urged providers to become familiar with food assistance programs and understand the role of social determinants of health (e.g., homelessness, poor literacy, low education, income) and tailor treatment to these issues.[1] Our health care system is becoming aware of how the conditions in which people live, learn, work, and play affect health.[20-32]

Household food insecurity is common, and T1D and T2D are increasing in YYAs

Food insecurity is at near-record-high levels in the US: In 2015 one in eight households was food insecure (12.7%).[14] One in six households with children was affected (16.6%).[14] In parallel, the SEARCH for Diabetes in Youth Study documented a 21% increase in T1D and a 31% increase in T2D prevalence in US youth between 2001 and 2009.[33] T1D incidence increased to a much greater degree in Hispanics and non-Hispanic blacks than in non-Hispanic whites (4.2% vs. 2.2% vs. 1.2%, respectively).[34, 35] T2D incidence has also increased (4.8% overall), particularly in non-Hispanic blacks (6.3%) and Hispanics (3.1%). These trends suggest that an increasing number of YYAs, particularly minorities and many with food insecurity, will be burdened with diabetes. Although T1D is generally associated with higher socioeconomic status (SES)[36] (whereas T2D is associated with lower SES),[37-39] our

preliminary data on YYAs show that food insecurity affects 19% of those with T1D and 38% of those with T2D, both higher percentages than the 2015 national prevalence of 12.7%. Moreover, food insecurity affects an increasing number of middle-income households.[40]

The adverse effects of household food insecurity are severe, lifelong, distinct from poverty, and disproportionately affect racial/ethnic minorities.

Household food insecurity in childhood is associated with poor subsequent mental and physical health, including depression, dysthymia, behavioral problems, and poor social skills in older children, as well as iron-deficiency anemia, lower bone density, poor reported health status, and poor physical functioning in young children.[41-53] Moreover, a multitude of studies have now shown that food insecurity's effects on health outcomes are distinct from those of SES and poverty, a finding confirmed among youth.[54-56] Given these associations, it is concerning that food insecurity disproportionately affects racial/ethnic minorities. Non-Hispanic blacks (21.5%) and Hispanics (19.1%) had a higher prevalence of household food insecurity in 2015 than non-Hispanic whites (10.0%).[14] Thus, household food insecurity may help explain outcome disparities among YYAs with diabetes and lead to interventions that address food insecurity and eliminate disparities.[16]

Household food insecurity is particularly challenging for persons with diabetes

Because diabetes treatment requires careful alignment of medication and dietary intake, persons with T1D and T2D face additional health challenges because of food insecurity. Unpredictable food availability leads to not only worry and anxiety but also skipping meals or not eating for long periods of time. In older adults with T2D, food insecurity was associated with more episodes of severe hypoglycemia, poorer glycemic control, more unfavorable lipid and blood pressure levels, more episodes of depression, and more emergency department (ED) visits and hospitalizations.[8, 11, 12, 57-63] Similar studies among YYAs with T1D or T2D are lacking.

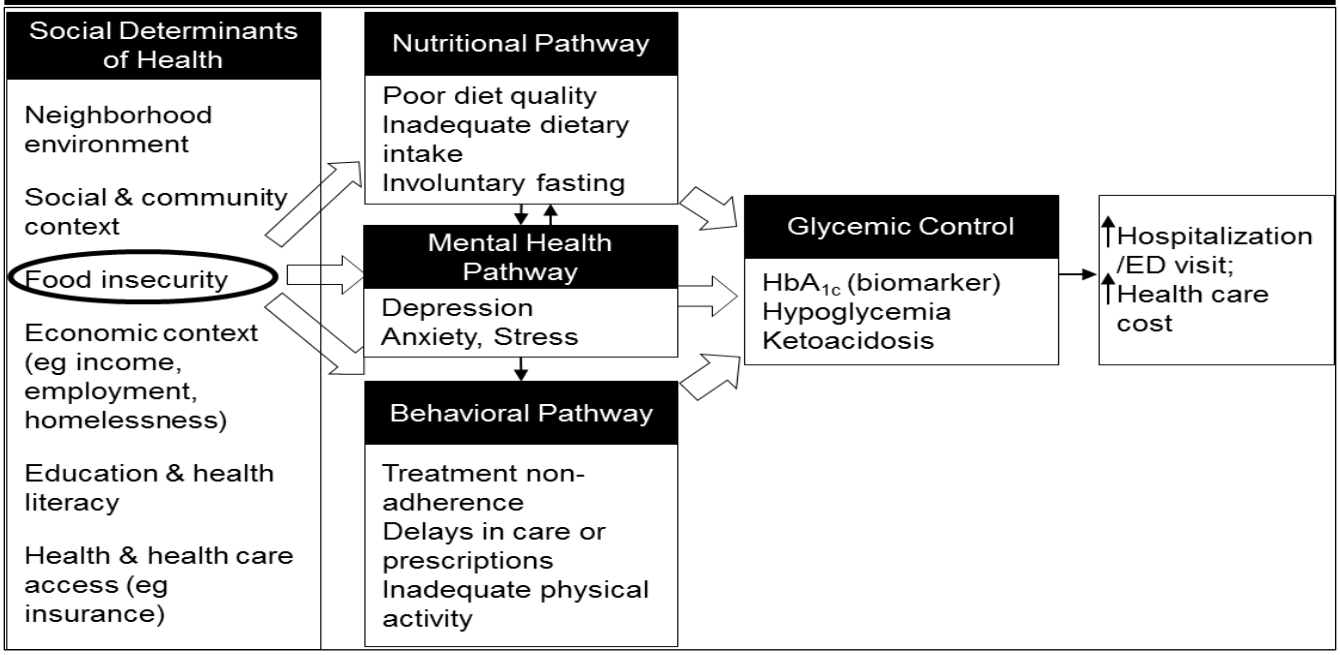
Household food insecurity acts through several pathways and diverts resources from diabetes self-management

A framework developed by Weiser et al. of how food insecurity affects health[64, 65] includes three main pathways: (1) nutritional (e.g., hunger, undernourishment), (2) mental health (e.g., depression, anxiety, stress), and (3) behavioral (e.g., poor coping strategies). ***Our conceptual framework (Figure 1) builds directly on Weiser's and has been adapted to reflect the unique situation of diabetes***. We posit that each of the three pathways leads to specific diabetes-related manifestations, which all divert resources away from diabetes self-management. This contention is supported by the general public health literature in that hunger and undernourishment are often associated with poor meal planning and disordered eating patterns among YYAs with T1D and T2D.[66-70] Likewise, depression, anxiety, and stress lead to delays in seeking care, filling prescriptions, and medication scrimping because of trade-offs between food and medicine purchasing (the "eat or treat dilemma" faced by 30% of food-insecure families).[8, 10, 58, 63, 71-73] Poor coping strategies are also related to poorer diabetes self-efficacy, more physical inactivity, and poorer diabetes self-managment.[74, 75] Each pathway may lead independently to poor glycemic control, but they can also influence each other. No study has formally quantified these pathways, which we propose to do by utilizing causal inference–based methods for mediation analyses described by VanderWeele.[76] ***This study does not focus on the causes of food insecurity but on the consequences of food insecurity and the mechanisms by which food insecurity may exert its effects in YYAs with diabetes.***[77-79]

Food assistance programs alleviate but do not solve household food insecurity

Federal food assistance programs such as the Supplemental Nutrition Assistance Program (SNAP) provide scheduled monthly benefits and are designed to cover ~70% of food expenditures;[80] thus, a sizeable proportion of recipients still experience episodes of food insecurity.[81] Furthermore, the SNAP program reaches ~83% of the eligible population.[82, 83] Although the effectiveness of food assistance programs in general populations has been evaluated,[84] their particular benefit in persons with diabetes has not been studied.

Figure 1 Conceptual framework of food insecurity's impact in persons with diabetes



Strengths and weaknesses of previous research

Since the 2014 review by Gucciardi et al.,[85] new papers have emerged, but the nature of the previous research is unchanged:[59, 61-63, 86-94]

(1) ***Most studies focused on middle-aged and older adults with T2D*** who had adverse health consequences associated with household food insecurity. Only one paper focused on youth with T1D or insulin-dependent T2D,[12] and none examined young adults with T1D or T2D.

(2) ***Virtually all previous studies have been cross-sectional***, which limits causal inferences. Only two previous studies have been prospective, both intervention studies in adults with T2D, and neither one included more than two time points, limiting the ability to fully tease out causal mechanisms.[60, 61] The Treatment Options for Type 2 Diabetes in Adolescents and Youth trial did not measure food insecurity.[95-98]

(3) ***Few studies have explored the role of mediators (pathways)*** through which food insecurity affects health outcomes, and none have used state-of-the-art statistical mediation approaches. The studies of adults with T2D by Berkowitz et al.[61] and Seligman et al.[8] offer important evidence in support of our mediation hypotheses related to nutritional, mental health, and behavioral pathways.[8] Both studies used conventional mediation analyses, which require strong assumptions that are often not met.[76]

(4) ***No studies have explored the impact of youth's own food insecurity experiences with regard to diabetes,*** which may have effects beyond the experience of household food insecurity.[99-102]

(5) ***Very few studies on glycemic control have considered food insecurity in the context of disparities and controlled for other social determinants of health*** — most have only controlled for education and income.[8, 11, 12, 71, 86, 87, 103] We propose to examine food insecurity as a mechanism for disparities in health outcomes among YYAs with diabetes,[3] providing a target for interventions for eliminating food insecurity and the disparities. If this approach appears promising, then further examination of mechanisms linking minority race/ethnicity and food insecurity will be investigated such as bias/discrimination, nativity/immigration status, neighborhood social cohesion, and food assistance enrollment.[104-107]

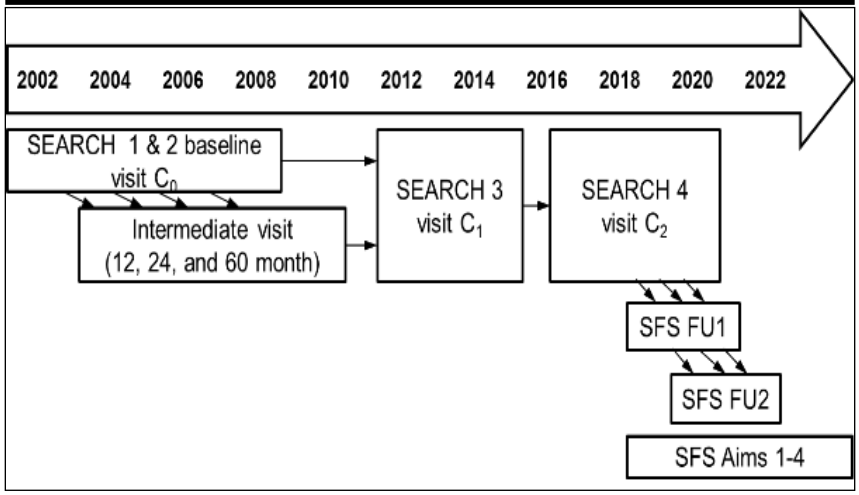
(6) ***No studies have evaluated the shape of the relationship between food insecurity and glycemic control.*** If we confirm our preliminary finding that the relationship between food insecurity and glycemic control is not

linear, these findings would have strong implications for practice guidelines, including a more in-depth food insecurity assessment for persons who are positive on the two-item screener, and more nuanced recommendations for care depending on level of food insecurity and HbA_{1c}.

Study Design

The proposed study will leverage the ongoing SEARCH for Diabetes in Youth study (SEARCH 4, 2015-2020) to conduct a food insecurity–focused longitudinal ancillary study, the ***SFS cohort study***, that prospectively examines the association of food insecurity with glycemic control, health care utilization, and costs in YYAs with T1D and T2D (**Figure 2**). We will integrate SEARCH data with new SFS data. The SFS design was guided by our conceptual framework (**Figure 1**) and developed to (1) comprehensively assess food insecurity in YYAs with diabetes; (2) optimize temporal clarity and separation between assessment of exposure (e.g., food insecurity), mediators (e.g., nutritional, mental health, diabetes self-management behaviors), and outcomes (e.g., glycemic control, health care utilization, medical and non-medical health care costs); (3) efficiently leverage the racial/ethnic diversity of SEARCH; and (4) include YYAs with T1D and T2D. The proposed work complements but does not duplicate SEARCH 4 or any SEARCH ancillary study and is supported by SEARCH.

Figure 2 SFS study design in relation to SEARCH



Overview of SEARCH for Diabetes in Youth study

The SEARCH parent study is a multi-center epidemiological study that ascertained incident cases of diabetes diagnosed at age <20 years starting in 2002 and continuing to the present.[16,46,65] With approval of local Institutional Review Boards, eligible cases were asked to complete an initial patient survey (IPS). For incident cases in 2002-2006 and 2008, individuals who completed the IPS were invited to participate in a first in-person research visit (IPV) where clinical and biochemical data were collected (response rate 71.5%). The sample participating in the baseline IPV was representative of all registered cases in terms of clinical and metabolic characteristics.[108] The SEARCH study has multiple components to address surveillance and etiologic research goals and is currently in its fourth cycle (SEARCH 4). ***For brevity, we describe only those elements within SEARCH that are directly relevant to the present proposal***.

Design of SFS cohort study

The SFS cohort study will utilize data from three time points (**Figure 2**), including SEARCH 4 plus two SFS data collections spaced nine months after SEARCH 4. To maintain and maximize efficiency while maintaining excellent racial/ethnic diversity and numbers of participants with T2D, the SFS study will be conducted at three of the five SEARCH sites: Colorado (CO), South Carolina (SC), and Washington (WA).

Table 1 presents the starting population for SFS (n=1,187; 556 white, 631 minority) by location and diabetes type and recruitment goals, anticipating 90% retention for follow-up 1 (FU 1 n=1,069; 501 white, 568

minority) and 2 (FU 2 n=962; 450 white, 512 minority). Revised estimates for participants eligible from SEARCH and recruitment success for SFS are presented in Table 1a. (see eligibility section for explanation). Data to be collected under the SFS cohort study are shown in **Table 2** and will include (1) repeat assessment of food security and food assistance at FU 1 and 2, (2) information on hypothesized mediators (e.g., nutrition, mental health, diabetes self-management behaviors), (3) information on covariates (confounders) characterizing other social determinants of health not assessed in SEARCH 4 (i.e., literacy, homelessness), and (4) outcomes (e.g., HbA_{1c}, hypoglycemia, ketoacidosis, health care utilization and costs). More detail will be provided in Section 4 Data Collection.

Because SEARCH 4 data collection started in July 2016, ~600 of the estimated 1,187 SFS-eligible persons will have 12-27 months of follow-up time between SEARCH 4 and SFS FU 1 (October 2018), the rest will have 9 months follow-up. SFS FU1 recruitment will be front-loaded, balancing rapid recruitment of those with longer follow-up times with maintaining the 9 months follow-up for the rest. FU 1 and 2 will be nine months apart. Thus, SFS FU 2 will be ~18-36 months after a participant's SEARCH 4 visit or completed surveys if only surveys were completed in SEARCH 4. *Due to the Coronavirus, the FU2 clinic visit component was suspended in March 2020, to protect the health and safety of research study personnel and research participants. Because of the ongoing coronavirus pandemic, HbA1c will be collected from participants via dried blood spot instead of a blood draw in a clinic setting. The dried blood spot method can be remotely self-administered by the participant and returned by mail. Participants who prefer to have the dried blood spot administered in a clinic setting will be offered that option, if available. SEARCH classification of diabetes type at recruitment uses provider-based information supplemented by diabetes autoantibodies (DAAs) and insulin sensitivity.[17]

Table 1. Estimated participation in SFS (90% retention)

Location	SEARCH 4 (age 15-35)			SFS FU 1 (age 15-37)			SFS FU 2 (age 16-38)		
	T1D	T2D	Total	T1D	T2D	Total	T1D	T2D	Total
CO	413	89	502	372	80	452	335	72	407
SC	283	150	433	255	135	390	229	122	351
WA	225	27	252	203	24	227	182	22	204
Total	921	266	1,187	830	239	1,069	747	215	962

Age range (years) at enrollment in SEARCH 1: 0-19 for T1D and 10-19 for T2D.

Table 1a. Revised estimated eligible and participation in SFS based on revised retention estimates (75%) and addition of survey only group

Location	SEARCH 4 (age 10-35)	SFS FU 1 (age 10-37)	SFS FU 2 (age 10-38)
	Total	Total	Total
CO	716	537	417
SC	551	413	324
WA	363	273	212
Total	1,630	1,223	953

Age range (years) at enrollment in SEARCH 1: 10-19 for T1D and 10-19 for T2D.

Eligibility Criteria

The SFS cohort builds on the sample recruited in SEARCH 4 (PI D'Agostino Jr, 1 UC4 DK108173-01) composed of persons with diabetes diagnosed before age 20 years in 2002-2006, 2008, and 2012 who had had a SEARCH baseline visit earlier in the study and were seen for an in-person visit during SEARCH 3 (**Figure 2**). SEARCH 4

Table 2. Data available from SEARCH 4 and measures for SFS follow-up (FU) 1 and FU 2. (Note survey data collection at FU will be conducted online and by mail; red denotes new measure)

Variables	SEARCH	SFS	
	4	FU1	FU2
Food security (household and child-specific) and assistance, date of benefit distribution	X	X	X
Socio-Economic variables (education, employment, income, health insurance)	X	X	X
Other social determinants of health (housing & homelessness, numeracy, discrimination, transportation)		X	X
Laboratory and anthropometric measures (HbA_{1c}, height, weight)	X		X
Episodes of self-reported hypoglycemia and ketoacidosis	X	X	X
Health care utilization	X	X	X
Health care cost	X	X	X
Demographics (age, sex, race/ethnicity)	X		
Clinical characteristics (diabetes type, diabetes duration)	X	X	X
Diabetes management (glucose monitoring, medication regimen, adherence)	X	X	X
Problem eating	X		X
Dietary intake (screener)		X	X
Depression	X	X	X
Anxiety		X	X
Perceived stress, perceived social support		X	X
Physical activity and inactivity	X	X	X

optimizes prior data and efficiently collects additional data needed to elucidate the impact of diabetes on the health of YYAs with diabetes. Thus, across all five centers, *recruitment targets all participants with T2D, all minority YYAs with T1D, and a random sample of non-Hispanic white YYAs with T1D, with a goal of n=1,846 for SEARCH 4. Note that SEARCH has always met its recruitment goals in the past, but it is now unclear if this will hold true for SEARCH 4.*

SEARCH has recently extended its data collection to 9/30/2019 in hopes of reaching the goal, as recruitment success has been lower than expected and it will definitely be lower than the original estimates set forth while planning SFS (see Table 1). Thus, SFS plans to extend recruitment to SEARCH 4 Survey-Only participants to offset lower than expected recruitment in SEARCH 4 cohort visit participants. Survey-Only participants were part of SEARCH 3 and complete the SEARCH 4 online surveys but do not participate in the SEARCH 4 in-person clinic visit procedures. The Survey-Only participants all have a diagnosis of type 1 diabetes and are of non-Hispanic white race/ethnicity. Thus, while additionally recruiting the Survey-only will allow SFS to meets its overall recruitment goals (see Table 1a), it will selectively add a sample of non-Hispanic white T1D to the study population. Table 1a shows new estimates for eligibility and participations in SFS under the assumption of 75% recruitment success and addition of the Survey-Only group to the recruitment pool.

The age range for SEARCH 4 is estimated to be 15-35 years for T1D and 18-35 years for T2D. ***The participants of SEARCH 4 will form the basis of the SFS study, as this time point is the first at which food security status was assessed on all participants SEARCH-wide.*** Key data of SEARCH and the SFS study are shown in **Table 2**. Eligibility for data collection will follow SEARCH 4's protocol, which reschedules pregnant women for data collection at 4 months after pregnancy.

Other eligibility criteria relate to when SEARCH participants become eligible for SFS FU 1 and FU 2 are as follows:

We define two groups at the point of SFS data collection initiation: a) Backlog group: S4 cohort visit and/or survey-only participants who are immediately eligible for SFS FU 1 on 11/1/2018, b) Ongoing group comprised of S4 cohort visit and/or survey-only participants who are becoming eligible incrementally. For the Backlog Group, because date of S4 visit and/or survey-only is 9 months or greater in the past, these individuals are eligible immediately for recruitment. Their S4 visit and/or survey-only date ranges from S4 inception to 11/30/17. For the Ongoing Group, S4 visits and/or survey-only have taken place/will take place between 12/1/17 and 9/30/2019. Last date for a SFS participant to become eligible for FU 1 is 6/30/2020. Data collection for FU 1 will continue until 12/31/2020 as a buffer period. For individuals who complete FU 1, eligibility for FU 2 starts on 6/1/2019, as does data collection, and continues through 12/31/2020. However, for individuals who do not complete FU 1 (Backlog or Ongoing groups) the eligibility window for FU 2 opens 4/1/2020 and closes 12/31/2020 and is followed by a 6-month buffer period through 6/30/2021. In all SFS recruitment priority will always be given to those participants that originated in the S4 cohort visit group, as they have the more complete data.

Recruitment and retention success in SEARCH 1-4.

SEARCH has been able to recruit and retain participants over time: 75-82% of SEARCH 1 and 2 participants were recruited to the SEARCH 3 cohort visit (82% in SC, 75% in CO, 75% in WA). However, SEARCH 3 was five years after the previous in-person visit, a much larger interval than what is proposed here. SEARCH 4 has a target 75% recruitment rate, but is currently not on track to achieve this goal. Retention over a shorter 12-month interval obtained from the SEARCH 2 12-24-month follow-up recruitment was 91% (90% T1D, 92%T2D; 96% white, 89% minority) but these past successes now seem optimistic. SFS will evaluate its recruitment success systematically about 6-7 months after study initiation to develop more precise recruitment estimates. At the current time, we believe a 75% recruitment success may be achievable for those participants in the SEARCH 4 cohort visit group, and at best a 75% recruitment success in the survey-only group. These statistics are the basis of our revised recruitment estimates as of 1-31-2019 (Table 1a).

The infrastructure of SEARCH at each site has remained consistent. SEARCH recruitment, retention, and tracking is facilitated through ongoing bi-annual mailed correspondence (birthday card and annual contact update form). In the contact update form, the participant can confirm or update current contact information either online or via mail in a prepaid envelope ($5 incentive for returning the form). SEARCH also utilizes medical records and online databases such as LexisNexis and a range of approaches and technologies (e.g., mail, email, telephone, and texting). Because of the tight linkage of SFS personnel and SEARCH staff, we are confident that we will achieve the same levels of recruitment and retention as SEARCH.

Data Collection

The SFS study will use the same rigorously standardized methods as the SEARCH protocol. New measures selected specifically for the SFS study, including homelessness, health literacy, anxiety, stress, and social support, are based on validated instruments. **Table 2** shows the timing of data collection. The section below describes all measures: (1) exposures, (2) outcomes, (3) mediators, (4) other covariates (confounders) and moderators. Relevant citations are provided in text and **Table 3** to facilitate evaluation of scientific rigor.

Logistically, the SFS FU 1 will be conducted via phone, mail and online. Participants will be contacted by SFS staff and invited to participate an online survey on a secure interface using a unique study ID created by the Coordinating Center (CoC). For participants without internet access, mail or phone-based surveys will be used and data will be entered by study staff.

SFS FU 2 will utilize the same approach as FU 1 with respect to the surveys. Additionally, the participants will be invited to attend an in-person clinic visit at which a non-fasting blood sample will be drawn for HbA_{1c} determination and their height and weight measured. *Due to the Coronavirus, the FU2 clinic visit component was

suspended in March 2020, to protect the health and safety of research study personnel and research participants. Because of the ongoing coronavirus pandemic, HbA1c will be collected from participants via dried blood spot instead of a blood draw in a clinic setting. The dried blood spot method can be remotely self-administered by the participant and returned by mail. Participants who prefer to have the dried blood spot administered in a clinic setting will be offered that option, if available. Results of the dried blood spot HbA1c will not be shared with participants or their providers. The DBS analysis is designed for research studies and not clinical purposes nor clinical decisions-making. * Due to this protocol change, height and weight measurements will be foregone as they are not essential to any stated hypothesis. If needed, height and weight can be estimated based on the participant's last SEARCH visit. The time window allowed between FU 2 survey completion and in-person FU 2 visit will be 3 months maximum. *Due to the Coronavirus, the FU2 clinic visit component has been suspended to protect the health and safety of research study personnel and research participants. The window between the FU2 survey and clinic visits will be allowed to exceed 3 months.*

An overview of the timing of the various data collection elements is presented in **Table 2.** Exposure measures: Food security.

Household food security

Household food security is the main exposure of interest and is ascertained using the 18-item HFSSM just as in SEARCH 4 Module 17. Parents/guardians of SEARCH participants under age 18 and participants with diabetes ≥18 years of age complete the HFSSM, which measures household food insecurity over the previous 12 months scored on a continuous linear scale ranging from 0 to 10 (no children in household) or 18 (children in the household), with higher scores indicating more severe food insecurity.[6] The US Department of Agriculture (USDA) provides equivalent scale values from 0 to 10 for households with and without children (the "standard 0 to 10 metric") to allow for direct comparisons using continuous HFSSM score data.[6] HFSSM's reliability has been reported as 0.86-0.93 using the Spearman-Brown estimate, Rulon's split-half estimate, and Cronbach's alpha.[109] Validity of the HFSSM has been established as summarized in a report by Dr. Frongillo (co-I).[110] The USDA's Guide also provides guidance on a series of skip patterns that are to be applied to reduce respondent burden. These skip patterns were not programmed for the SEARCH 4 Module 17 online interfaces, nor are they represented on the SEARCH 4 paper versions. For SFS, the skip patterns have been programmed and will be indicated clearly on the paper forms, following updated guidance issued in September 2012.

Child-reported food security

Child-reported food security among participants 10–17 years of age is ascertained using the six-item CFSA questionnaire developed and validated by Dr. Frongillo et al. at the University of South Carolina (see Appendix).[99, 100] The six items map to four domains of child food insecurity: cognitive awareness, emotional awareness (two items), physical awareness (two items), and responsibility/initiation. Youth report how frequently they have experienced each aspect of food insecurity over the last year, with responses of "never," "1 or 2 times," and "many times" coded as 0, 1, and 2. Each item has good-to-excellent accuracy as an indicator of its intended domain.[100] Child food security will be utilized in Aim 3, sub-aim 3.1, where it will be added to a model already including household food security.

Table 3. SFS study variables		
	Description	Refs
Exposure		
Household food security (HFSSM)	USDA household questionnaire	6
Child reported food security (CFSA)	Child reported experiences of food security	99, 100
Outcomes		
Glycemic control	HbA_{1c}	3, 111, 112
Hypoglycemia & Ketoacidosis	Parent- or self-reported episodes	113
Health care utilization	Hospitalizations, ED visits, urgent care, diabetes & primary care clinics	113
Health care cost	Medical and non-medical health care costs via MEPS and US Bureau of Labor Statistics data	114-116
Mediators		
Dietary intake	Diet Screener Questionnaire DSQ in NHIS 2015 Cancer Control Supplement	117
Depression, anxiety, stress	CES-D, GAD-7, Cohen's stress scale	118-121
Diabetes self-management behaviors	Glucose monitoring, type of device, adherence to Rx regimen	113
Problem eating	SEARCH Diabetes Eating Problem Survey (DEPS-R)	
Physical activity & inactivity	IPAQ	122, 123
Perceived social support	Social support scale	124, 125
Resilience	Connor-Davidson Resilience Scale 10 (CD-RISC-10)	
Covariates/Confounders		
Demographic data	Age, sex, race/ethnicity, and marital status	113
Duration of diabetes	Via date of initial diabetes diagnosis	113
Insulin regimen	Pump +/- injections with insulin type, basal & bolus dosages & frequency	126
Other medication regimen	Any other meds, including Metformin	113
Tobacco use	Current tobacco use	113
Socio-economic variables	Education, employment, income (includes federal and state benefits), household composition	113
Health insurance status	Public, private, other, none	113
Housing / Homelessness	Adapted from DHHS definition	127-129
Diabetes numeracy	Diabetes Numeracy Test-5, Subjective numeracy scale	130,164,165
Transportation	Type of transportation and reliability	
Discrimination	Everyday experiences of discrimination and particular events	131
Moderator		
Race/Ethnicity	White vs. Minority race/ethnicity	113
Diabetes type	T1D vs. T2D	17, 113
Food assistance	SNAP, WIC, food banks, & free/reduced school lunch program	132-135

Outcome measures

All outcomes will be assessed at SFS FU 2 on the SFS cohort.

HbA_{1c} and glycemic control

HbA_{1c} is the standard measure of glycemic control over the past three months for people with diabetes and is the primary continuous outcome for the proposed study. In the SFS cohort study, HbA_{1c} was initially measured following the SEARCH protocol and utilizing the same central laboratory, the Northwest Lipids Laboratory (PI: Dr. Marcovina). HbA_{1c} was measured in a sample of whole blood taken from participants during an in-person visit. The sample is analyzed with an automated nonporous ion-exchange high-performance liquid chromatography system (model G-7; Tosoh Bioscience, Montgomeryville, Pennsylvania).[3] The intra-assay coefficient of variation is 0.047%, the inter-assay coefficient of variation is 0.070%, and the normal reference range values are 4.2–5.8%.[113]

Due to the Coronavirus, the in-person clinic visits for the whole blood draw was suspended in March 2020, to protect the health and safety of research study personnel and research participants. In addition, the Northwest Lipids Laboratory closed and is no longer an option for HbA1c analysis.

As of July 2020, HbA1c will be collected from participants via dried blood spot (DBS) utilizing the University of Washington School of Medicine Department of Laboratory Medicine (PIs: Dr. Potter; Dr. Wener). The laboratory will perform the processing (punching) the DBS samples, elution of DBS punches for analyses, and measure the percentage of glycosylated hemoglobin (A1c) in DBS eluates, and return the test results to the study team. DBS methodology is well established and has been validated in numerous studies against HbA1c determined on a blood sample obtained via venipuncture (Lacher et al. 2014; Crimmins 2014)

Glycemic control will also be categorized using the ADA and International Society for Pediatric and Adolescent Diabetes (ISPAD) 2014 Guidelines for HbA_{1c} as follows: for ages <18 years, (1) <7.5% is optimal, (2) 7.5-9.0% is suboptimal, and (3) >9.0% is high risk;[111, 112] for ages ≥18 years, (4) <7.0% is optimal, (5) 7.0-9.0% is suboptimal, and (6) >9.0% is high risk.[3, 111, 112] HbA_{1c} will be measured at SFS FU 2.

Hypoglycemia and diabetic ketoacidosis

Participants will also be queried on frequency of severe hypoglycemia and diabetic ketoacidosis (DKA) using the SEARCH questionnaires Module 6 at SFS FU 2.[113]

We considered alternative strategies for measurement of these constructs in response to reviewer comments but maintain that self-report of severe hypoglycemia and DKA is the most efficient approach to collecting these data: First, because diabetic ketoacidosis (DKA) is a serious medical event requiring emergency room treatment or hospitalization, self-report of ketoacidosis is very accurate, which is why SEARCH ascertains self-report of DKA per its protocol. For severe hypoglycemia, defined as episodes in which the assistance of others was needed, very high agreement between prospective reporting of hypoglycemia every month compared to long-term recall over the past 12 months (89.6% agreement) has been reported among adults with type 1 diabetes.[136] A recent study reported similar rates of severe hypoglycemia comparing prospective monthly assessment versus 6-month long-term recall of hypoglycemia.[137] Alternatives to self-report would include downloads from glucose meters if available, but those would be subject to the same biases and limitations as described below. Thus, we will likely miss episodes of non-severe hypoglycemia.

Health care utilization

SEARCH 4 queries the (1) type and frequency of health care provider visits (for diabetes and primary care); (2) number of hospitalizations for one or more nights; (3) ED visits; and (4) visits to an urgent care facility in the last 12 months, as well as the number of diabetes-specific utilizations in Module 6.[113-115] Health care utilization and costs are assessed with an online survey in SEARCH and SFS FU 1 and FU 2.

Medical and non-medical health care cost.

SEARCH 4 queries medical and non-medical health care costs in Module 9 and Module 10. Module 9 will be kept in its entirety because it contains relevant information on health insurance and out of pocket expenses. Module 10 will be kept in its entirety. We will assess these modules at FU 1 and FU 2.

Health care costs will be assessed using a micro-costing approach previously applied by Dr. Wright in other settings.[116, 138, 139] Total costs are estimated by multiplying units of resource utilization by a unit cost for that resource, e.g., for medical costs related to health care visits, we multiply the number of provider visits by the cost of each visit. The unit cost for face-to-face urgent care, ED, and primary care visits and the cost of an inpatient stay will be derived from data on health care expenditures in the most recent Medical Expenditure Panel Survey (MEPS),[140] which provides nationally representative estimates of health care utilization, costs, and payment sources. We will calculate mean and median costs associated with a service for all MEPS participants age 10-35 years for baseline cost assessment and for age 10-38 years for follow-up assessment using the total expenditure variable, including costs paid by insurance, out of pocket, or by other sources. In addition to medical costs, we will estimate non-medical costs of health care, including costs related to missed work time and childcare.

SEARCH 4 queries how much time participants (or their parents) have taken off work or school to attend a medical appointment or because of a health problem over the previous 12 months. Lost productivity will be valued at the mean hourly earnings of the US full-time civilian workforce obtained from the US Bureau of Labor Statistics ($23.23 in 2015).[141,142] Respondents also reported the number of times they had to make childcare arrangements to attend a medical appointment for their diabetes and the average number of hours required. Total childcare costs will be valued at the average household productivity wage rate.[143] All costs will be inflated to common year dollars using the Consumer Price Index. Health care expenditures estimated from self-reported utilization and non-medical expenditures will be summed to estimate total costs. We will construct 95% confidence intervals around costs using non-parametric bootstrapping techniques. Because SEARCH 4 also queries health care expenditures for the preceding six months in detail, including categorical information about health insurance structure and coverage, cost of premiums, deductibles, co-pay amounts, and monthly out-of-pocket expenditures for different types of health care visits and services,[144, 145] the SFS study will also be able to conduct analyses on self-reported expenditures.

Mediators

Mental health, nutritional, and behavioral mediators will be assessed at SFS FU 1 and 2 via online questionnaires, supplemented by mailings or phone calls as needed. The mediating pathways (Aim 3) have been directly informed by the conceptual framework by Weiser et al. (Figure 1).[64, 65] Note the mental health assessments will not be conducted in children/youth age 10-13.

Depressive symptoms

Depressive symptoms will be assessed using the CES-D[118], a widely used 20-item questionnaire adapted for children and adolescents.[146] In adolescents, a score of ≥24 is suggestive of depression[119] and warrants further psychological evaluation, whereas a score of 16 or higher has been used for adults.[118] We will use the three-category, 24-cut-point stratification developed by Rushton et al[119] to assess depression severity in adolescents (minimal (0–15), mild (16–23), and moderate/severe (24–60)) and use a score of ≥16 for adults. High scores on the CES-D will trigger immediate online information of resources and study contact info and prompt follow-up by study staff. Youth under 18 who have high scores will be notified immediately online, study staff will also reach out to the participant's parent or guardian to notify them as well. The original 20-item CES-D scale[118] does not include questions on suicidal ideation. A revised CES-D scale was created in 2004[147], which included two questions on suicidal ideation. The SEARCH study has used the original scale, and thus this scale, without assessment of suicidal ideation, will be used in SFS.

Anxiety (new measure)

We will use the seven-item Generalized Anxiety Disorder (GAD-7) screener by Spitzer et al.[120, 148] The GAD-7 is well established as an online tool and has been applied in youth and adults with diabetes.[149-151] Each item is rated on a likert-type scale from zero ("not at all sure") to three ("nearly every day"), with total scores ranging from 0 to 21 (< 5 minimal anxiety; 5-9 mild anxiety 10-14 moderate anxiety; 15+ severe anxiety). Individuals with scores of 10+ are recommended for further assessment (Spitzer et al, 2006). High scores on the GAD-7 > 15 will trigger study contact info and prompt follow-up by study staff who will provide materials and resources. For child

and adolescent participants, their parent/guardian will also be notified by study staff. We made one revision to a GAD-7 answer category indicating the lowest level of agreement with each question: Previously worded "Not at all sure" was changed to "Not at all" to improve clarity and comprehension. This revised wording has been used by many other investigators since the original publication of this screener. See the following search for examples:

https://www.google.com/search?q=GAD-7&client=firefox-b-1-ab&source=lnms&tbm=isch&sa=X&ved=0ahUKEwjAzPDbnKTdAhXK3VMKHUeyDwIQ_AUICigB&biw=1280&bih=887#imgrc=TdWH5-yIar9NBM:

Perceived stress (new measure)

We will use the validated 14-item Cohen's Perceived Stress Scale, which measures the degree to which situations in one's life are considered stressful on a five-point Likert scale.[121] The Perceived Stress Scale had high levels of internal consistency (Cronbach's alpha=0.86) in our previous research on food security.[128, 152]

Perceived social support (new measure)

SFS will additionally introduce a measure of perceived social support, as social support has been an important factor associated with food security in previous work[153] and is thought to buffer against the impact of stress according to Cohen's stress-buffering hypothesis.[124] Social support will be measured using the 12-item Multidimensional Scale of Perceived Social Support,[124, 125] which measures perceived support from family, friends, and a significant other on a seven-point Likert scale. The perceived social support scale had high levels of internal consistency (Cronbach's alpha=0.93) in our previous work, including in food-insecure and food-secure populations.[128, 154]

Other behaviors and factors relating to diabetes self-management, such as glucose monitoring, medication non-adherence, physical activity and inactivity will be measured following the SEARCH 4 protocol.

Resilience (new measure)

SFS will additionally introduce a measure of resilience in the second follow-up survey. The 10-item Connor-Davidson Resilience scale (CD-RISC) includes 10 items rated on a 5-point scale with higher scores reflecting greater resilience (score range: 0-40) (Connor and Davidson 2003). Per agreement with the developers of the scale, all future users of the data from this scale must refer to the CD-RISC agreement and stipulations.

Dietary intake (new measure)

Participants' dietary intake will be assessed with the 29-item Diet Screener Questionnaire applied in the NHIS 2015 Cancer Control Supplement study. (NCI DSQ 2018) It is nearly identical to the DSQ administered in the NHANES 2009-10, "with the only difference being that in 2015 two questions were asked to capture intakes of sports drinks, energy drinks and fruitades, whereas only one question was asked in the NHANES 2009-10 DSQ."[155] Used in conjunction with a scoring algorithm, the DSQ allows derivation of predicted intake of fruits, vegetable (with and without excluding French fries) including legumes, and fiber, dairy, calcium, added sugar, sugars from sugar sweetened beverages, and whole grain. We made a very small wording modification to this instrument on the question pertaining to beans. Original wording was "During the past month, how often did you eat refried beans, baked beans, beans in soup, pork and beans or any other type of cooked dried beans? Do not include green beans." This was changed to: "…refried beans, baked beans, beans in soup, pork and beans or any other type of canned or cooked dried beans? Do not include green beans."

Diabetes self-management

SFS will ascertain diabetes self-management characteristics exactly as in SEARCH 4 (Module 3 and 4), including (1) glucose monitoring attributes such as the type of glucose measurement device used (glucometer vs. continuous glucose monitor), (2) frequency of blood glucose measurement with the glucose meter, (3) type of actions taken as a result of actual glucose values, and (4) medication regimen (including the use of continuous subcutaneous insulin

infusion, CSII) using standardized questions.[113, 126] The frequency of self-monitoring of blood glucose has been associated with improved glycemic control in patients with T1D.[111, 112] Additionally (5) Insulin non-adherence is assessed by asking how often and why participants missed taking their insulin; (6) diabetes medication adherence (e.g., Metformin) is similarly assessed. SEARCH Module 5 which includes more peripheral information on diabetes management will not be included in SFS.

Physical activity and inactivity

SEARCH 4 characterizes physical activity and inactivity using the International Physical Activity Questionnaire short form (IPAQ-SF) and Marshall's sitting questionnaire.[122, 156] The IPAQ-SF consists of seven items that assess vigorous- and moderate-intensity physical activities, as well as walking and sitting. The IPAQ-SF had acceptable reliability in US adults (Spearman's rho=0.66-0.88).[157] Likewise, reliability for total sitting time and for attaining >150 min/week of moderate-intensity or greater physical activity had acceptable reliability (rho=0.71-0.94) and percent agreement (0.86-1.0).[158] Validity with accelerometry had acceptable percent agreement for attaining recommended levels of physical acivity.[159, 160] The entire SEARCH 4 Module 19 consisting of 6 questions will be used in SFS.

Problem eating

In follow-up 2, SFS will ascertain disordered eating in all participants consistent with SEARCH 4 using the Diabetes Eating Problem Survey (DEPS-R). Disordered eating can negatively impact glycemic control for individuals with diabetes. The 16-item diabetes-specific screening tool has high levels of internal consistency (Cronbach's alpha=0.86) in a previous study (Markowitz, 2010).

Covariates (confounders) and moderators

Covariates acting as confounders or moderators have been selected as shown in (Table 3). Moderators include minority race/ethnicity and diabetes type. In addition, receipt of food assistance will be considered a potential buffer and evaluated as a moderator. Other social determinants of health will also be assessed in this category of measured variables, including homelessness, health literacy and discrimination, transportation, health insurance status.

Age

Age is assessed by self-report, querying the date of birth of the participant on the SEARCH IPS form and computing the age relative to the date of the data collected.

Sex

Sex is assessed by self-report on the participant's IPS form – Adult Version / Parent Version

Race/ethnicity

Sex is assessed by self-report on the participant's IPS form – Adult Version / Parent Version

Household composition, living alone, going to college etc.

Household composition among participants is assessed by self-report on the SEARCH 4 Survey Packet (Module 16) and entails a question asking "What is your current living situation?" Questions also include "How many people are currently living in your primary household all or most of the time including yourself?" and "Do you live or stay in more than one home on a regular basis?" Having children is assessed among SEARCH 4 adult participants in Module 16, question 9, which is being modified to assess the birth dates of children in order to be able to characterize changes in household composition over time, which may be related to changes in food security status.

Marital status

Marital status is assessed in the household questionnaire (Module 16) among adult and young adult respondents and parents of teens with a simple question on whether the respondent is married, divorced separated, never married, a member of an unmarried couple, doesn't know or prefers not to answer.

Education

Education among adult participants is assessed by self-report on the SEARCH 4 Survey Packet (Module 16) and entails a question asking for the highest degree or level of school the participant/spouse, father and mother has completed. 17 answer categories ranging from "No schooling completed" to "Professional or doctorate degree (for example, MD, DDS, JD, PhD, EdD)." "Don't know is also an option. Teens are asked to self-report on the highest degree of schooling they have completed. Parents of minor participants provide self-report of parent/guardian and the other parent/guardian if two adults in household.

Attending college is assessed among adult participants in Module 16, question 10. On the question on highest degree of schooling completed we modified the SEARCH 4 response category labeled "Nursery school to 4th grade" to read "Preschool to 4th grade."

Employment

Employment status among participants is assessed by self-report on the SEARCH 4 Survey Packet (Module 16) and entails a question asking for current employment status for at least 10 hours. Eight answer categories -1) Employed full time, 2) Employed part time, 3) Unemployed, seeking work, 4) Unemployed, not seeking work, 5) Student, not seeking work, 6) Disabled, 7) Other (Specify), 8) Don't know – are included in the survey.

Health insurance status

Health insurance status among adult participants is assessed by self-report on the SEARCH Healthcare Usage Form (Module 9) and entails 5 questions around health insurance status. The first question asks "In the past 12 months, were you covered by a health insurance plan or did you have a state or federally funded source of coverage for your healthcare?" Answers include yes, no, and don't know/ refused. A subsequent question asks "Who provided your health insurance?" Answer options include Medicaid, Medicare, CHIP, private, military, Indian health services, other, and don't know/refused.

Health insurance related to minor participants is asked of their parent/guardian following a similar structure focused on the kind of health insurance plans.

Income

Income among participants is assessed by self-report on the SEARCH 4 Survey Packet (Module 16) and entails a question asking for income received from personal earnings before taxes for adult participants and questions asking for total income in household for adult participants and parent participants. This includes wages or salaries, including tips, bonuses, and overtime pay, and income from self-employment. 14 answer categories range from "Less than $5000" to "$150,000 or more" and contain varying numbers of answer categories and width across questionnaires. "Don't know" and "Prefer not to answer" are also options.

Tobacco exposure

Tobacco use will be assessed using most but not all of the SEARCH 4 questions in Module 20, specifically the 10 specific tobacco-use questions, but not the 4 environmental tobacco exposure questions.

Alcohol and drug use

Alcohol and drug use will be assessed with 8 questions and 5 questions respectively, following the SEARCH 4 protocol on Module 20 and this will be assessed on teens age 10-17 (in SFS 10-17) and all adults.

Diabetes type

Diabetes types is based on the provider-based diabetes type consistent with SEARCH and categorized as type 1, type 2 and other.[161] We also have the capacity to use etiologically determined type.[17]

Food assistance

Food assistance is queried in SEARCH with 4 questions (Module 17), all framed within the past 12 month and for the entire household, asking about SNAP and WIC benefits and whether the participant or any household member receives emergency food assistance from a church, food pantry, food bank or eats in a soup kitchen, and lastly asks about the respondent's child receiving free or reduced price lunch at school. Response options are yes, no, refuse, don't know. The SFS study will additionally ask for the date of food assistance distribution.

Other federal assistance program participation

Receipt of any federal assistance, including welfare, Medicaid, and food stamps is queried in SEARCH 4 Module 16, question 14 for adult participants, but not for parents of minors. In SFS, this question will be asked of all participants, including parents of minors, where the wording will be changed to focus on the time frame relevant to the SEARCH participant's life, not the entirely life of the parent.

Number of pregnancies and outcomes

The number of pregnancies and pregnancy outcome and date thereof will be queried using SEARCH Module 8 question 4 at FU 1 and 2. This information may serve as a control variable in sex-stratified analyses of health care cost.

Social determinants of health

Transportation (new measure)

Transportation will be assessed using modifications of a questionnaire previously used in the Midlands Family Study[162] and asks 4 questions and queries having reliable transportation in the past 12 months, mode of transportation, access to transportation, and financial assistance for transportation to health care visits. The answer categories for primary mode of transportation were modified to include more recent transportation options (e.g. car share, ride share). The question on transportation to medical appointments was added de novo. The answer categories for primary mode of transportation for non-emergency medical appointments list examples of transportation services offered at each of the 3 sites (Colorado, Seattle, South Carolina).

Numeracy (new measure)

Numeracy will be assessed using an objective measure and a subjective measure. Numeracy has been associated with diabetes control.[163] Diabetes numeracy (objective measure) will be assessed using the 5-item Diabetes Numeracy Test.[164] This is a self-administered test that allows assessment of typical scenarios involving computations or assessments needed for diabetes-self-management. The 5-item short form was developed from the original 15-item form. [130] For the SFS study, this instrument will be adapted from paper and pencil administration to online administration, which will require small changes in wording. Interpretation of the DNT-5 results should always refer back to the normative studies done on this instrument, because these questionnaires test knowledge and are intended to be administered in a controlled setting, which we cannot guarantee in an online administration. Thus, additionally, numeracy will be assessed using the 3-item Subjective Numeracy Scale, the validity of which has been evaluated in persons with and without diabetes.[165] These two instruments will be placed towards the end of the questionnaire set.

Racial discrimination (new measure)

Racial discrimination will be assessed using the Everyday Discrimination Scale developed by Williams et al. which captures more "chronic, routine, and relatively minor experiences of unfair treatment" using 9 items.[131]

Diabetes discrimination

SEARCH 4 already assesses discrimination due to diabetes in Module 23 and this questionnaire will be retained for SFS at FU 1 and 2. Children and teens age 10+ and adults are asked to respond.

Housing and homelessness (new measure)

Housing situation/homelessness will be assessed using questions that capture the current DHHS (Administration for Children and Families) position on homelessness. These include questions about having stable housing and where people have slept in the last 90 days with 6 very specific answer categories and two questions on whether they were at risk in terms of health or safety or whether there is violence or conflict at the place where they are staying.[166] For some of the potential responses on question 1, examples from other documents were added including from "Frequently asked questions about health care for the homeless by the national health care for the homeless council"[167] and "Changes in the HUD Definition of "Homeless"" by the National Alliance to End Homelessness.[168]

Participant addresses

At the request of NIDDK, participant residential addresses will be collected in a systematic fashion to support potential future geographic information system-based work in the SFS study. A data collection form will be created to abstract the residential address at the time of the SEARCH 4 visit from the respective tracking databases and used again for address at SFS FU 1 and 2. Addresses will be stored locally until these are integrated into a GIS project.

Overview of data collection modules relative to respondent type and age

Table 4: Data collection modules in SFS relative to age group and respondent type

		Respondent in SFS				
Module	SFS Status	Adult 26+	Young Adult 18-25	Parent of 10-17 participant	Teen 14-17	Child 10-13
1 - Intro	Revise	YES	YES	YES	YES	YES
2 – Name & contact	Keep for FU1 and FU 2	YES	YES	YES	YES (short)	No
3 – Diab manage beh	Keep for FU 1 and FU 2	YES	YES	YES	YES	---
4 – Blood sugar manage	Keep for FU 1 and FU 2	YES	YES	YES	YES	YES
5 – more manage	Drop for FU 1 and FU 2	YES	YES	YES	YES	YES
6 – Hypo/DKA * HC Util	Keep for FU 1 and FU 2	YES	YES	YES	---	---
7 – Medical history	Drop for FU1 only.	YES	YES	YES	---	---
8 – Reproductive history	Keep only modified question 4 for FU 1 and FU 2 (see file below).	YES	YES	NO	YES	YES
9 – Med and non-med costs	Keep only questions: 1 2 3 4 7 8 9 10 for FU 1 & FU 2	YES	YES	YES	---	---
10 – Time and finance costs	Keep for FU 1 and FU 2	YES	YES	YES	---	---
11-15 - NA		---	---	---	---	---
16 – education/income	Keep for FU 1 and FU 2	YES	YES	YES	YES 1 question only –	YES 1 question only –

	Include modified question 9 on having children Include 1 question on marital status for adults and parents (same as on parent form) Include question 13 on public assistance on adult and parent forms.				highest level of school completed	highest level of school completed
17 - food assistance & security	Keep for FU 1 and 2 Insert 1 new question on date of benefit receipt into adult/parent module	YES	YES	YES	YES, youth version	YES, youth version
18 - NA		---	---	---	---	---
19 - PA	Keep at FU 1 & FU 2	YES	YES	---	YES	YES
20 – Tobacco/alcohol/drugs	Keep questions 1-10, 15-27 for FU 1 and FU 2	YES	YES	---	YES	YES
21-24 - NA		---	---	---		
25 – diab discrimination	Keep at FU 1 and FU 2	YES	YES	---	YES	YES
CES-D*	Keep at FU 1 and FU 2	YES	YES	---	YES	---
FFQ	Drop					
MNSI Form	Drop					
Physical Examination Form	revisit for FU 2					
Specimen Collection Form	revisit for FU 2					
Eating problem survey	Keep for FU2 only	YES	YES	---	YES	YES
SFS Unique measures						
Diet quality screener	New	YES	YES	---	YES	YES
Transportation	New	YES	YES	YES	YES	---
Housing & homelessness	New	YES	YES	YES (dropped violence and conflict question)	YES (dropped violence and conflict question)	NO
Discrimination	New	YES	YES	---	YES	YES
Numeracy (DNT-5 + SNS-3)	New	YES	YES	---	YES	YES
Anxiety	New	YES	YES	---	YES	NO
Stress	New	YES	YES	---	YES	NO
Social support	New	YES	YES	---	YES	NO
Resilience CD-RISC 10 item	New	YES	YES		YES	NO

Study Timeline

Figure 3. Timeline of the SFS study

Calendar Year	2017			2018				2019				2020				2021				2022				23
Quarter	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1
Study Year					Year 1				Year 2				Year 3				Year 4				Year 5			
Timeline of parent study (SEARCH 4) (started September 2015)																								
Timeline of proposed SFS cohort study																								
Finalizing study protocol, IRB review, hire staff																								
Staff training																								
Data collection SFS FU 1																								
Data collection SFS FU 2																								
Data management and cleaning																								
Statistical analyses																								
Presentations and publications																								

Data Management

Data Management and Quality Control

Data management will follow the SEARCH model. Each center retains identifying information. Participants will complete online surveys on a secure interface using a unique study ID created by the Coordinating Center (CoC). Mail or phone-based data from participants without internet access will be entered by study staff. The CoC manages the data and will routinely provide information on survey completion. Analysis datasets will be generated and shared with investigators at intervals. Laboratory data on HbA_{1c} will be sent from the University of Washington Laboratory to the CoC and from there to the data collection sites. DBS results will not be distributed to participants.

Sample Size and Power

Because all statistical analyses will be conducted within the framework of generalized linear models (GLMs, described below), power considerations can be examined jointly by considering (1) sample sizes available for different hypotheses and (2) potential associations of confounding variables and mediating variables with the outcomes. We anticipate having a cohort of 747 T1D and 215 T2D YYAs (total 962, 450 White, 512 minority race/ethnicity) for the longitudinal analyses (**Table 1)**. All power calculations used PASS version 13.[169] For the time being these estimates remain unrevised.

Power for GLMs relies on significance tests of regression coefficients, which are equivalent to tests of correlation (e.g., between HFSSM score and HbA_{1c}). For ***Aims 2 and 4, using a 0.05 two-sided test of the null hypothesis that the Pearson correlation coefficient is 0.0, we will have 90% power to detect a correlation of 0.104 or larger for 962 YYAs, 0.151 for 450 non-Hispanic white YYAs, 0.143 for 512 minority YYAs, 0.118 for 747 YYAs with T1D, or 0.219 for 215 YYAs with T2D.***

Table 4. Detectable correlations with 90% power and alpha=0.05 (2 sided) for Aim 3 mediation analyses

Partial R^2	Sample Size	Detectable R^2	Detectable R
.10	215	0.0423	0.206
.25	215	0.0353	0.188
.5	215	0.0235	0.153
.10	747	0.0125	0.112
.25	747	0.0104	0.102
.5	747	0.0070	0.084
.10	962	0.0097	0.098
.25	962	0.0081	0.090
.5	962	0.0054	0.073

For power considerations for the Aim 3 hypotheses on the mediating effects of variables on the relationship between the HFSSM score and an outcome of interest (i.e., HbA_{1c}), we must specify the amount of explained variance (R^2) of the outcome that the mediators account for. Second, we must specify the amount of explained variance that the key predictor (HFSSM score) will predict for the outcome, conditional on the mediators in the model. Finally, we must specify the available sample sizes.[169] **Table 4** shows scenarios for the three most extreme values of aforementioned sample sizes, partial R^2 values from the mediators (we conservatively assume up to 10 mediators), and detectable R^2 from the predictor of interest. Even with a sample size of 215, there is 90% power to detect an R^2 of 0.042 attributed to HFSSM score using an F-test with alpha=0.05 (two-sided), assuming that the R^2 explained by the mediating variables is 0.10. Thus, this study has excellent statistical power to detect associations of HFSSM score with outcomes of interest (HbA_{1c}, health care utilization and costs) and in fact has sufficient power to allow for analyses to be performed stratified by diabetes type (T1D/T2D). This is also true for sub-aim 3.1 (role of child-reported food security), for which we anticipate a sample of 373 youth ages 15-17 years with longitudinal data, of whom 335 will have T1D.

Statistical Analysis

Statistical analyses for Aim 2.

We will consider both parametric and non-parametric methods to test the association between household food insecurity (measured with HFSSM score) and change in glycemic control (outcome variable: HbA_{1c}).

Because participants were not randomized to food-secure or -insecure conditions, other characteristics could confound the associations of interest. We will use a propensity score method to estimate the conditional probability (propensity score) that a participant is food insecure based on their background covariates: age, sex, race/ethnicity, diabetes duration, insulin regimen, prescription medications, body mass index, tobacco use, household composition, education, income, employment, health insurance, homelessness, and health literacy. Note we are using a non-matching based approach, which is important given our sample size. We will first dichotomize household food insecurity into the USDA-defined food-secure and -insecure categories.[6] We will then estimate a propensity score (using logistic regression) for each participant and use this score in subsequent models to balance the food-secure and -insecure participants on their background characteristics. We will group participants into propensity score quintiles and compare each of the background covariates in the food-secure/-insecure groups, adjusting for the propensity score quintile and food-secure/-insecure group by propensity score quintile interaction. If this interaction is non-significant, we will remove it from the model and compare groups to determine whether balance on background covariates is achieved after adjustment for propensity score quintile. If balance is not achieved in the first iteration, a more complex propensity score model will be fit that includes interactions and/or higher-order terms of the covariates. Dr. D'Agostino Jr. has extensive experience in propensity score methods.[170-181]

The statistical analyses for Aim 2 will use GLMs for the relationship of household food insecurity (HFSSM score, continuous) with outcomes (e.g., HbA_{1c}, continuous), adjusting for propensity score quintile. As noted in our preliminary data, the relationship between HFSSM and HbA_{1c} may be not be linear. Therefore, as we fit the model, we will include a quadratic term for HFSSM (HFSSM squared), as well as a binary indicator variable for HFSSM score=0 (yes/no). One difference between this model and that used in the preliminary data is that here, we will examine longitudinal change in HbA_{1c} as it relates to change in HFSSM.

Using a longitudinal mixed model, we will consider the participants as random effects and HFSSM score, propensity score quintile, and time as fixed effects. Because the timing of visits differs by individual and all individuals have diabetes, the variable duration of diabetes (months) will be used as the measure of time. Each participant will have measures of time at their first visit or completed surveys (SEARCH 4 visit/surveys), SFS FU 1 (~9-27 months after SEARCH 4), and SFS FU 2 (~18-36 months after SEARCH 4) linked to their date of diabetes diagnosis. In addition to these fixed covariates, minority race/ethnicity (Aim 2, hypothesis 2) diabetes type (aim 2, hypothesis 3) and food assistance (Aim 2, hypothesis 4) will be considered as fixed covariates. Some variables will

be time-varying (duration of diabetes, HFSSM) and others (propensity score quintile and diabetes) will be fixed and non–time varying. We will first test for an interaction of HFSSM with propensity score quintile to determine whether the relationship between HFSSM and the outcomes is consistent across the spectrum of background characteristics (summarized in the propensity score quintiles). If this interaction is non-significant, as expected, we will remove the interaction term and fit the model. If the interaction is significant, we will fit the GLM within each propensity score quintile.

We propose to fit a sequence of models that become progressively more complex. Our first model will test directly for association between HbA_{1c} and HFSSM, adjusting only for the propensity score quintile, duration of diabetes, and diabetes type. Here we will also examine whether higher-order terms for HFSSM (or an indicator variable for HFSSM) are needed. Next, we will consider additional adjustments in the model, including covariates that were included in the propensity score model: age, gender, race/ethnicity, education, income, etc. and other social determinants of health (Aim 2, hypothesis 1). Whereas in the propensity score model, these variables were included because they represented characteristics of the participants that may have predicted whether their households were food secure or insecure, the inclusion of these variables in the longitudinal mixed model will add precision because of their potential association with the outcome (i.e., HbA_{1c}). To further assess the impact of food insecurity on glycemic control, we will fit a similar series of models as described above for HbA_{1c} on the related outcomes of number of episodes of hypoglycemia and diabetic ketoacidosis, as these are two clinical manifestations of poor glycemic control. Additionally, these outcomes can be considered as binary (e.g., onset of hypoglycemia among those without hypoglycemia at baseline) in the framework of a multiple logistic regression model. For all models, we will examine the fit of the data to the distributional assumptions, including homogeneity of variance, conditional normality, influence diagnostics, and outliers. If the assumptions of the linear model are not met and there are no appropriate transformations to allow for a linear model to be fit, we will consider fitting a non-parametric model. We will use an unstructured covariance structure for the longitudinal mixed models.[182]

To address Aim 2, hypothesis 2, the comparison of minority race/ethnicity in these models, main effects and race/ethnicity by HFSSM interactions will be included. These interactions will be with both time-varying HFSSM (measured at SFS FU 1 and 2) and "baseline" HFSSM (treated as a fixed effect) measured at SEARCH 4. By examining both of these interactions (time-varying and fixed), we can determine a differential impact on HbA_{1c} for racial/ethnic minority and non-Hispanic white YYAs depending on starting HFSSM and change in HFSSM. Although the outcome HbA_{1c} is assessed at two time points (SEARCH 4 and FU 2), HFSSM is available at three time points (SEARCH 4, FU 1, FU 2), and thus this additional information about time-varying HFSSM can be incorporated into the statistical models. If the interaction is significant, we will stratify models by major race/ethnicity categories to conduct subgroup analyses (white 450, minority 512). To address Aim 2, hypothesis 3, the impact of diabetes type in these models, we will similarly examine main effects and include a diabetes type by HFSSM interaction, and we anticipate conducting subgroup analyses by diabetes type. Note, two-way interactions between minority race and type cannot be explored given the sample size.

To address Aim 2, hypothesis 4, moderation of the association by food assistance, food assistance will be included as a main effect and an interaction to determine a potentially differential impact on HbA_{1c} depending on participants receiving food assistance. We will then select a subsample of low-income participants, i.e., household annual incomes <$50,000, and examine whether food assistance moderates the association between food insecurity and change in HbA_{1c}. This income threshold closely approximates two times the federal poverty threshold for a family of four and indicates eligibility for many federal programs.[183]

Statistical analyses for Aim 3.

The purpose of Aim 3 is to quantify the mediating role of nutritional, mental health, and behavioral pathways through which food insecurity may affect changes in glycemic control. We hypothesize that household food insecurity affects changes in HbA_{1c} directly and indirectly through nutrition, mental health, and diabetes self-management behaviors. The SFS study has been designed specifically to guarantee the correct temporal ordering of variables needed for mediation analyses.[184]

We will use the same longitudinal mixed models and propensity score methods as described for Aim 2 but will additionally use the counterfactual approach to mediation methods described by VanderWeele[76] to evaluate multiple mediators at the same time. This approach relies on less-stringent assumptions than the conventionally used method of Baron and Kenny.[185] It also allows for effect decomposition of the total effect of food insecurity into a direct effect and multiple, pathway-specific indirect effects, even in the presence of interactions and nonlinearities. Thus, this method allows assessment of the relative contributions of the food-insecurity pathways operating through

the mediators (nutrition, mental health, diabetes self-management) and not through the mediators (i.e., directly) and thereby quantification of the importance of these pathways.

Analyses for sub-aim 3.1 will be restricted to 373 youth age 10-17 with longitudinal data (335 with T1D), as they are eligible to complete the CFSA. This sample will be reduced by possibly up to 22 individuals who are 10-13 being excluded from the analyses focusing on mental health because those measures will not be assessed in this age group. In this 10-17 year old group, we will perform analyses evaluating the role of child-reported food insecurity on changes in HbA_{1c}, independent of household food insecurity, using the same sequence of longitudinal mixed models as for Aim 2. We will then expand these models to include the complexities of the above-described mediating pathways. We will compare the partial R^2 values that include additional information from the different models examined, including the additional variability explained with the addition of the CFSA to the overall models. Using partial F-tests, we will determine if the additional information from the CFSA adds significantly to the model.

Statistical analyses for Aim 4.

To prospectively evaluate household food insecurity in relation to changes in health care utilization and costs in YYAs with T1D and T2D, we can use the same type of GLM and propensity score method as described for Aim 2 because the propensity score model was not linked to a specific outcome but rather to the exposure (HFSSM). Additional analyses could focus on whether different patterns of HFSSM (e.g., improving, stable, or worsening) predict different health care costs. We will also examine whether there is a differential impact on cost based on race/ethnicity or diabetes type by testing the corresponding interactions (Aim 4, hypothesis 2 and 3). Analyses could be stratified by race/ethnicity or diabetes type and models re-fit for race/ethnicity or T1D and T2D separately. For Aim 4, hypothesis 4, we will build on the GLM described above and the tests for interaction with food assistance and then select a subsample of low-income participants, i.e., household annual incomes <$50,000, and repeat the analyses. From this sample, we will examine whether food assistance moderates the association between food insecurity and change in health care utilization and costs.

Challenges and Strengths

Design and analysis.

The prospective longitudinal design overcomes cross-sectional design limitations; propensity score methods and marginal structural models will improve causal inferences. We will avoid all connotation of causality, however, because this study is not a randomized trial.

Threats to validity.

Self-selection to the study poses risks to validity.[108] The SEARCH 3 participants have similar distributions of demographic, clinical, and socioeconomic variables as those who were registered cases eligible for the study but did not participate. In SFS, we retain the ability to evaluate selective participation because we have data on everyone from previous visits. Thus, we believe that the SFS study will have equally high internal and external validity as the SEARCH study.

Exposure assessment: applicability or specificity of HFSSM to persons with diabetes.

Even though the food security questionnaires (e.g., HFSSM, CFSA) were not developed specifically for persons with diabetes, these instruments are already being used in SEARCH, well validated, and widely used, allowing for comparisons to other studies and national estimates. Moreover, the ADA screening recommendations include two questions from the HFSSM questionnaire, which increases the relevance for people living with diabetes.

Outcome assessment: lack of daily blood glucose data.

Although the SFS will collect HbA_{1c}, it will not collect glucose levels from continuous glucose monitoring (CGM), insulin pump data, finger-stick measurements, or meter downloads. CGM is the state-of-the-art measure for daily blood glucose assessment, but the cost of such a protocol would exceed this application's budget. We considered including meter downloads or other non–systematically collected data on blood glucose, but these can be

highly biased, as patients who are more compliant with glucose monitoring are also more likely to have better glycemic control. Thus, this study will not have the capacity to examine associations between the date of monthly disbursement of SNAP benefits and daily blood glucose control or other research questions linked to short time intervals.

Health care utilization and cost data: self-reported measures.

Although these data are subject to recall error, major events such as hospitalizations have acceptable validity (kappa=0.89). Although ED and office visits are underreported by 33% and 19%, respectively, underreporting was not clinically different across demographic groups (nonsystematic), and thus results would be biased towards the null.[186]

Missing data.

We will determine if the missing pattern is ignorable using the quadratic inference functions of Qu and Song.[187] If it is, we will apply traditional methods for missing at random patterns; if informative, more sophisticated statistical methods will be applied, such as modeling conditionally on the pattern of missing data.[179-181] For missing HFSSM data, we will apply USDA-designed imputation methods.[6]

Food security and other social determinants of health.

This study will assess food insecurity as one of four social determinants of health (income, education, and literacy) and is thereby broad in scope to address questions of interrelationships between them. However, assessment of neighborhood environment and social/community context is beyond the scope of this study, as the resources needed to characterize the neighborhood environment at a state-of-the-art level exceed the capabilities of this study.

Summary of strengths.

First, this proposal leverages the size and racial/ethnic diversity inherent in SEARCH 4 as the largest and most diverse cohort of YYAs with T1D and T2D ever assembled, with 53% minority YYAs. Second, it captures an inception cohort, with all participants having laboratory-confirmed T1D or T2D along with substantial amounts of longitudinal data from SEARCH. Third, it integrates data on biological, behavioral, sociocultural, and health care factors assessed by standardized, validated methods. Fourth, the SFS study's design has been informed directly by an existing, well-developed conceptual framework[64, 65] explicating the mechanisms by which food insecurity affects health and testing the proposed pathways via nutrition, mental health, and diabetes self-care behaviors. Finally, the SFS cohort study substantially expands the scope of the SEARCH study to address important questions on how food insecurity longitudinally affects glycemic control and health care utilization and costs among YYAs with diabetes with the highest levels of scientific rigor, thereby shedding light onto existing disparities in diabetes management and control.

Human Subjects

Protection of Human Subjects

As required by the parent SEARCH study policies and procedures, the review by SEARCH investigators of the resubmission application entitled "Food Insecurity, Glycemic Control, and Hospitalizations among Youth with Diabetes" to which we will henceforth refer as the SEARCH Food Security (SFS) cohort study included careful consideration of additional burden that this ancillary study would place on SEARCH participants. It was determined by the SEARCH Ancillary Study Committee that the proposed study would impose some additional burden on SEARCH participants and that this additional burden was appropriate relative to the new knowledge that would be generated. Furthermore, it was determined that the additional data collection required by the proposed study would not adversely impact on the likelihood of ongoing participation in SEARCH. All aspects of research done under the auspices of the parent SEARCH project have been approved by local institutional review boards and all activities have been deemed HIPAA compliant. *Here, we focus only on the additional activities relevant to the protection of human subjects that would occur as a result of participation in the proposed SFS study.*

This Human Subjects Research meets the definition of 'Clinical Research'.

Risks to Human Subjects

Human Subjects Involvement, Characteristics, and Design

As described in the Research Design and Methods, this study will involve the follow-up of participants of the ongoing SEARCH 4 cohort visit and/or survey (2015-2020) funded by NIH/NIDDK (UC4 DK108173, PI D'Agostino Jr.) all of whom were diagnosed with diabetes between 2002 and 2012 and were less than 20 years of age at the time of diagnosis.

The SEARCH subject population defines the population eligible to be recruited into the SFS study. This population includes people who were diagnosed with diabetes when less than 20 years of age. Individuals who had gestational diabetes mellitus (GDM) only, who were active duty military at the time of diagnosis, and individuals who were institutionalized were not eligible for inclusion in the incident case sample. SEARCH has five clinical centers, located in Ohio, Washington, Colorado, South Carolina, and California. Cases of diabetes were identified in geographically-defined populations in Ohio (8 urban-suburban counties encompassing and surrounding Cincinnati); Washington (5 urban counties encompassing and surrounding Seattle); and Colorado and South Carolina (all counties in these states), among health plan enrollees in California (Kaiser Permanente Southern California), and among indicant health service beneficiaries in several American Indian populations.

The SFS study will be conducted at three of the five SEARCH sites, in the interest of maintaining scientific rigor and maximizing efficiency while maintaining excellent racial/ethnic diversity and numbers of participants with T2D: Colorado (CO), South Carolina (SC) and Washington (WA) and will recruit all SEARCH 4 cohort participants who participate in the SEARCH 4 clinic visit and/or surveys only, including T1D and T2D at those locations. Given the high levels of recruitment success (80% and more) experienced in SEARCH. SEARCH's follow-up periods range from 12 month to 5 years, and SEARCH has documented 91% retention rates over 12 month. Thus, the SFS study is anticipating a 90% recruitment and retention rate relative to SEARCH, given that it will begin recruiting within a 9-27 month time interval after SEARCH. Tables in the Research Strategy section of the application present the starting population for the SFS cohort study (n=1,187) by location and diabetes type and recruitment goals, anticipating 90% retention for the first SFS follow-up (FU 1 n=1,069) and the second SFS follow-up (FU 2 n=962) and the data to be collected under the SFS cohort study protocol.

No subjects will be excluded on the basis of gender or race/ethnicity. Because the focus of this study is to learn more about the impact of diabetes on people who are less than 20 years of age at the time of diagnosis, this study includes adolescents and young adults ranging from 10 to 35 years of age at the initiation of the SFS study. This study does not involve fetuses, neonates, prisoners, or institutionalized individuals. Young women, who are pregnant and eligible for the study, will not be invited to participate in Hba1c DBS collection until at least four months after the completion of their pregnancy

The study population will consist of approximately 53% females. The race/ethnic distribution is described in the planned enrollment table and includes 53% minorities.

All dried blood spot samples will be collected and sent to the University of Washington Department of Laboratory Medicine for analysis. The Coordinating Center will be responsible for data management and analysis. The laboratory is the University of Washington School of Medicine Department of Laboratory Medicine; and the Coordinating Center (CoC) for this study is Wake Forest University in Winston-Salem, North Carolina. The Coordinating Center serve in the same capacity for the SEARCH 4 study and the SFS study.

Sources of Materials

The SFS study will recruit all eligible subjects (described above) to participate in two (2) follow-up (FU) assessments in the context of the SFS study. The assessment will include the following research material:

1. Contact information (all participants): name of subject and parents or guardians (if <18 years of age) and alternate contacts, addresses, phone numbers, and e-mail addresses

2. Surveys and questionnaires (all participants) at FU 1 administered online or by mail/phone: Questions related to food security and food assistance, other social determinants of health, demographics, employment, education, income, clinical characteristics, behavioral diabetes management indicators, dietary intake, physical activity and inactivity, depression, anxiety, perceived stress and perceived social support.

3. Surveys and questionnaires (all participants) at FU 2 administered online or by mail/phone: Questions related to food security and food assistance, other social determinants of health, demographics, employment, education, income, clinical characteristics, diabetes management indicators, health care utilization, medical and non-medical health care costs.

4. Remote self-administered (or in-person, if available) assessments at FU 2:

 blood specimens (non-fasting): HbA1c

All of the data listed above will be collected both manually and electronically. Each research subject was assigned a unique SEARCH identification number at the time that he or she was originally registered for the study. The SFS study will continue to utilize the SEARCH unique identification number so that data can be linked with SEARCH data as required by this protocol. The data elements included in the FU 1 and FU 2 will be collected specifically for this research study.

Each of the SEARCH centers maintains subject names and contact information locally, accessible only to the local research team. The Protected Health Information (PHI) that is transmitted to the SEARCH Coordinating Center (CoC) for registered cases is the minimum necessary to conduct this research. It consists of date of birth, county, zip code, date of diagnosis for diabetes, and dates of inpatient and outpatient visits. Data transmitted to the CoC qualifies as a HIPAA Limited Dataset. Each of the centers has entered into a Limited Data Use Agreement with the CoC in compliance with the Standards of Privacy of Individually Identifiable Health Information as outlined by HIPAA. Local access to subject identifiers will be governed by the requirements of the local IRB. Laboratory specimens will be associated only with the SEARCH identification number and the date of specimen collection.

Potential Risks

The finger stick may cause a small amount of pain and bruising at the injection site and may bleed slightly. There is a rare chance that the stick site could become infected. This finger stick is similar to the one used by youth and young adults with diabetes when they monitor their blood glucose levels at home multiple times a day.
The household food security assessment, the assessment of children's self-reported food security status, receipt of food assistance, and health care cost information could lead to embarrassment and shame on the part of the participant or their parent/guardian, as there is some stigma associated with food insecurity in US society as well as privacy concerns for health care cost. Because the participant or parent/guardian will be completing the survey questionnaire on household food security status at a location of their choosing, a location that allows them to either access the online website or allows them to fill out the questionnaire on paper from their home, the risk of public disclosure and embarrassment is considered minimal.

The Centers for Epidemiologic Study of Depression (CES-D) is a scale that is designed to identify subjects who may be at increased risk for clinical depression and the Generalized Anxiety Disorder (GAD) Scale by Spitzer et al. is used to identify individuals at increased risk of anxiety. Although a revised version of the CES-D includes questions assessing suicidal ideation, the original version used in this study does not include an assessment of suicidal ideation. When a subject or their parent/guardian (if <18 years of age at the visit) is informed that either the CES-D or the GAD score is elevated, this knowledge may create additional anxiety. Each center has a protocol in place to deal with individuals with elevated scores. Given that the CES-D and GAD questionnaires will be administered online, an automatic scoring algorithm will be programmed by the CoC survey platform. Elevated scores will trigger a pop-up screen that provides contact information for staff and resources for depression and anxiety specific for each study site, as well as SFS investigators who can be contacted if more information is desired. In addition, an elevated score will trigger a notification of the site's project manager who will follow up with the study participant and/or their parent or guardian if under 18 by phone to offer further services such as referrals to a local clinical social

worker or other local mental health resources. Children age 10-14 will not be asked to complete the CES-D or GAD of the stress or social support scales.

Disclosure of information about medical conditions has the potential to embarrass, distress, or directly harm individuals.

Data collection carries risk of loss of privacy and confidentiality for individuals and for their parents. This research study does not include any treatment or intervention. The only alternative to participation in this study is to refuse participation. This decision will not impact the subject's medical care, insurance coverage, or health care/ treatment in any way.

Adequacy of Protection against Risks

Recruitment and Informed Consent

Subjects who are eligible for participation in the SFS study will be identified by the Coordinating Center. ***The SFS study will be restricted to those SEARCH participants who have indicated that they are interested in being informed about SEARCH ancillary studies.*** A letter will be mailed to these potential subjects from the local sites that gives a brief description of the SFS study and the follow-up assessments. If the subject is less than 18 years of age at the time of recruitment for the cohort visit, this letter will be mailed to their parent or guardian. Letters sent to subjects who are 18 years of age or older will be addressed to the subject. For the majority of eligible participants, SEARCH staff will already have informed them about the opportunity to participate in the SFS study while the subjects or their parents are participating in the SEARCH 4 study in-person visit. A designated member of the local research team will call the parent/subject to further describe the SFS study, to answer any questions the parent/subject might have, and to send the parent/subject the website link and login key and password for the FU 1 surveys. Upon opening the survey website, participants will be presented with the above mentioned fact sheet. Clicking a button indicates that they have read and consented to completing the survey and that will lead them to the start of the survey. Completion of the online survey forms will be considered active consent.

For FU 2, conducted 9 months after FU 1, the research team member will send another letter describing the second part of the study, send the new website link and login key and password. In addition, the research team member will coordinate the mailing of the dried blood spot remote self-administered blood collection (or in-person DBS, if available). DBS will be coordinated within 3 months of the survey completion. Subjects/parents who agree to participate in the DBS collection will be mailed the needed supplies and instructions and study staff will review the study requirements with the subject and/or parent and address any questions or concerns they might have. When the DBS kit arrives to the subject by mail (or in-person administration, if available) informed consent and assent forms for participants under 18 years of age will be included in the package. Assent of subjects who are less than 18 years of age is also governed by the requirements of the local IRB. If the subject is 18 years of age or older, the subject must give written informed consent. The informed consent will be documented in REDcap via secure e-consenting procedures prior to the packet being mailed. Copies of completed consent forms will be maintained in the subject's research record, according to local protocol.

Protections against Risk

To minimize the risk of loss of privacy, in-person DBS administration, if available, are conducted in a private location so that participant feels comfortable by trained staff. In the rare instances that a participant chose to have the DBS collected at a clinic site location, to minimize the possibility of risks associated with the finger stick in an in-person setting, experienced medical staff will obtain the blood samples.

Study personnel will be trained to identify the signs and symptoms of a blood glucose level that is low. They will also be trained to check the blood glucose level, using a glucometer. If a low blood glucose occurs (< 70 mg./dl.), study personnel will be trained to administer 15 grams of an oral carbohydrate, and to repeat as needed every 10 minutes until the blood glucose level is > 70 mg./dl. If the blood glucose level is above 300 mg./dl., study personnel will be trained to check urinary ketones.

On the CES-D questionnaire if a participant has a high score (> 24 if less than 18 years of age; > 16 if 18 years of age or older), or on the GAD questionnaire, a score of 15 or higher, the participant will see (parent will be contacted

if participant is < 18 years of age) a pop-up screen on the website listing a variety of mental health resources. . An automatic scoring algorithm will be programmed by the CoC survey platform. Elevated scores will trigger a pop-up screen that provides contact information for staff and resources for depression and anxiety specific for each study site. In addition, an elevated score will trigger a notification sent to the site's project manager who will depending on each site's approach, follow up with the study participant (or parent/guardian if participant is under age 18) by phone to offer further services such as referrals to a local clinical social worker or other local mental health resources.

After completion of the food security questionnaire, each participant will also see a link to a set of site-specific food assistance resources. Participants will also be given a site-specific housing resources sheet.

Data obtained during the visit will be recorded, both manually and electronically. The data management system for this study will utilize the combination of a local tracking application and a web browser-based interface. The local tracking application is a tracking database. It will be used by local study personnel to manage demographic data, contact information, consent, appointments, visits, and communications with the subject. This database will be password-protected and accessible to local study personnel only. The web browser-based interface will be used for recording the majority of the data collected as part of this study. Usernames and passwords will be required to access the SEARCH web site. The Coordinating Center will control web access rights by assigning individual usernames and passwords to each staff member, according to the level of access required. The web-based data entry system will protect confidentiality and data security by utilizing 128-bit encryption and Secure Socket Layer (SSL).

All Protected Health Information (PHI) will be used or disclosed in compliance with the Health Insurance Portability and Accountability Act (HIPAA). A limited amount of Personal Health Information (PHI) will be shared with the SEARCH Coordinating Center. This data includes month and year of birth, county, and, month and year of diagnosis for diabetes. Each of the centers have entered into data use agreements with the Coordinating Center in compliance with the Standards of Privacy as specified by HIPAA contingent on the interpretations and processes defined by the local IRBs/Privacy Boards. Local access to subject identifiers will be governed by the requirements of the local IRB.

Once all data have been collected for the SFS study, the Coordinating Center will generate data analysis data sets that will be shared with University of South Carolina investigators for statistical analyses for publications and presentations. These data sets will be de-identified in the sense that they will no longer contain identifying information in that the date of birth will be replaced by age at visit (in years) and the unique participant identifiers used by the SEARCH study, but a new identifier that can only be linked to the SEARCH identifier by staff of the Coordinating Center, but not by anyone else. This will provide an added layer of protection against disclosure as statistical analysts and student trainees at the University of South Carolina or at other institutions will not be in a position to identify any of the SFS study participants.

This study has been granted an automatic Certificate of Confidentiality from the National Institutes of Health. The researchers can use this Certificate to legally refuse to share information that may identify participants in any federal, state, or local civil, criminal, administrative, legislative, or other proceedings, for example, if there is a court subpoena. The Certificate of Confidentiality however does not release the investigators from the obligation under the law to report to the state suspected cases of child abuse or neglect, or suspected cases of intended suicide or intention to harm others.

Potential Benefits of the Proposed Research to Human Subjects and Others

1. Participants may receive a report with the results of their in-person clinic blood draw test (HbA1c) once the samples have been analyzed and sent to the local clinical centers by the University of Washington Laboratory. Laboratory results may also be shared with their healthcare provider with the participant's consent. In some cases, based on SEARCH test results, the healthcare provider may choose to make changes to the treatment plan. Participants who provide blood samples using the dried blood spot method will not receive a copy of their test results, nor will their health care provider. In addition, study participants/parents will receive an incentive of $50 for completion of the online surveys in FU 1, $ 50 for the completion of the online surveys for FU 2 with an added $10 increase to $60 if the survey is completed

within 2 weeks of the invitation, and $30 for the clinic visit (which included blood draw and height and weight measurement through March 2020; Dried blood spot samples only beginning in July 2020) for FU 2. Starting in December 2021, the dried blood spot incentive will be increased to $50 if returned within one month of receiving the offer. The incentives will be mailed to the participants as gift cards after they complete the online forms for FU 1. For FU 2, the participants will receive $80 (or $90 if the survey is completed within 2 weeks of the invitation) in gift cards when they provide a blood sample. If they do not provide a blood sample, they will receive the $50 (or $60 if the survey is completed within 2 weeks of the invitation) gift card via mail 3 months after completion of the survey, because that is the maximal allowed time window between these two data collection elements.

While all ancillary study participants will receive a standard document describing food assistance programs in their local area and state, there are no immediate, direct benefits of the proposed research to the subjects who participate in the proposed study. The main benefit of this study is that it will provide novel insight into whether food insecurity is associated with glycemic control, health care utilization and health care cost among youth and young adults with diabetes and potential mechanisms through dietary intake, mental health, physical activity, and glucose monitoring and self-management. The study will include an ethnically diverse population of youth with diabetes, which will advance current knowledge about the risks associated with diabetes in youth and provide guidance for food policy interventions.

Participation in this study may result in potential benefits to society. This is a large, multi-center study that will be well-represented by young people from a variety of racial/ethnic backgrounds. The information obtained in this study will help clinicians to better understand food insecurity patterns experienced by YYA with diabetes over time, and how food insecurity affects glycemic control, health care utilization and medical and non-medical health care cost. Overall, the information obtained by this study will fundamentally characterize the impact that food insecurity has on the lives of YYA with diabetes and has the potential to inform future revisions of the standards of care for persons with diabetes and the current food insecurity screening guidelines. This information will also be important in the planning of the distribution of medical and financial resources for the care of young people with diabetes in the future.

Potential risks to study participants are minimal and reasonable in relation to the anticipated benefits to society from the knowledge that will be gained from this study.

Importance of the Knowledge to be Gained

This study has major implications for clinical practice, policy, and future research. Findings of our study may provide evidence that (1) the current practice of using a two-question food security screening approach may need to be supplemented with a more in-depth assessment of food insecurity among those that screen positive so that a provider can tailor self-management advice to each particular patient (for instance, non-food days need to be treated like sick days, with appropriate reduction in medications). (2) The anticipated findings on pathways by which food insecurity affect health may suggest a new strategies for referring patients to federal, state, and local resources, including mental health support and nutritional support in the health system (e.g., dietitians) and in the community (e.g., food banks). (3) Last but not least we hope that our study will lead to findings translatable to reductions of health disparities among race/ethnic minorities. All of the above would affect future revisions of the ADA and AAP clinical practice guidelines. Our findings would also lay the groundwork for advocacy efforts for food assistance programs, both local/state and federal, which may need tailoring for persons with diabetes.

Last but not least, diabetes is the third most common chronic disease of childhood and adolescence. In the past, childhood diabetes was thought to consist almost exclusively of Type 1 diabetes. Over the past two decades, however, an increasing number of cases of Type 2 diabetes have been reported within this population. Overall, the total number of diabetes cases affecting people less than 20 years of age seems to be increasing over time. This is a large, multi-center study that will be well-represented by young people from a variety of racial/ethnic backgrounds.

Potential risks to study participants are minimal and reasonable in relation to the importance of the knowledge that is expected to be gained from this study.

Study Organization and Funding

Multiple PI Leadership Plan

The Food Insecurity, Glycemic Control, and Hospitalizations among Youth with Diabetes Study will have two principal investigators: Drs. Angela Liese and Jason Mendoza will both serve as the Principal Investigators for this grant application. Dr. Liese brings expertise in the epidemiology of pediatric diabetes, food insecurity research, epidemiologic methods and design and analysis of cohort studies, and in successfully conducting multi-center studies. Dr. Liese has also been involved in the SEARCH for Diabetes Study since its inception, which is important because the present application, the SEARCH Food Security (SFS) study, builds on SEARCH. Dr. Mendoza brings expertise in pediatric and community medicine, food insecurity research, epidemiology and cohort studies, and in behavioral and policy interventions to reduce nutrition and physical activity inequities among racial/ethnic minorities.

As a team, the multiple PIs (mPI) will provide oversight of the entire Research Strategy and all Specific Aims. They will develop and implement the policies, procedures, and processes for the Research Strategy to ensure uniform and high-quality data verification and processing all three proposed data collection sites: Carolina, Washington, and Colorado. Dr. Angela Liese is based at the University of South Carolina and will provide oversight and supervision of the study and staff for the Carolina site. Dr. Jason Mendoza is based at the University of Washington and the Seattle Children's Research Institute and will provide oversight and supervision of the study and staff for the Washington site. Together, they will also provide additional guidance to the data collection at the Colorado site led by Dr. Sauder, and hold weekly meetings with Dr. Sauder to plan and discuss data collection at the three sites.

The proposed study will be governed by a Steering Committee, co-chaired by mPI Drs. Liese and Mendoza. Other Steering Committee members will include Drs. Frongillo, Flory and Merchant at the University of South Carolina, Dr. D'Agostino Jr. and Reboussin at the Wake Forest University SEARCH Coordinating Center, Dr. Sauder and Dr. Dabelea at the University of Colorado, and Drs. Pihoker and Wright at the University of Washington/Seattle Children's Hospital. The Steering Committee will provide scientific leadership and oversight for the study. The Steering Committee will communicate monthly throughout the project to monitor progress and engage in scientific discussion. In addition, Drs. Liese and Mendoza, along with the project coordinators, will meet approximately weekly by phone to monitor progress. Decisions are made by consensus, with formal approval by the Steering Committee required for any modification to the protocol, and for papers using data. Formal approval is established by majority vote of the Steering Committee.

Herein we briefly describe the members of the steering committee—see biosketches for complete details. Dr. D'Agostino Jr, PhD (co-I), is the PI of the SEARCH Coordinating Center at Wake Forest University, the PI of the SEARCH 4 cohort study, and a professor of public health science. He is a national expert in applied statistics and has a long-standing track record in the areas of cardiovascular disease, diabetes, cancer, genetics, and statistical methodology, including extensive experience using propensity score methods to analyze observational data. Dr. Pihoker, MD (co-I), is a professor of pediatrics at UW and chief of the Division of Endocrinology and Diabetes at Seattle Children's Hospital. She has been the PI for the WA site since the inception of SEARCH and is one of only two practicing pediatric endocrinologists on the leadership team of SEARCH 4. Her extensive knowledge and experience in providing medical care for YYAs with diabetes will be important in the interpretation of the SEARCH clinical data and dissemination of results. Dr. Frongillo, PhD (co-I), is a professor in the Department of Health Promotion, Education, and Behavior at UofSC and an international expert on nutrition policies, household food insecurity, and hunger. Together with colleagues, he developed the CFSA and served as a member of the team that developed the HFSSM. Dr. Beth Reboussin, PhD, (co-I), is a professor of biostatistics and an expert in growth curve and latent class analyses modeling. Dr. Merchant, ScD, MPH, DMD (co-I), is a professor of epidemiology and biostatistics at UofSC. He has experience in causal modeling, particularly in using advanced mediation analyses, which he will apply in Aim 3 analyses. Dr. Sauder, PhD is a behavioral scientist who will lead the CO site. Dr. Wright is a highly trained health economist. In addition, four other investigators will provide guidance to the steering committee and mPIs because of their particular roles related to oversight of data collection in the SEARCH 4 study, including Drs. Mayer-Davis (SEARCH PI of Carolina site, University of North Carolina, coordination of SC SEARCH activities), Dr. Apperson (PI of SC Greenville site, pediatric endocrinologist), Dr. Bowlby (PI of SC Charleston site, pediatric endocrinologist) and Dr. Dabelea (SEARCH PI of CO site).

The study database will be developed and housed at the SEARCH Coordinating Center at Wake Forest University under the supervision and oversight of Drs. D'Agostino Jr (co-I) and Reboussin (co-I), with guidance from the mPI Drs. Liese and Mendoza. Statistical analyses will be conducted primarily at the University of South Carolina by advanced doctoral students in epidemiology or biostatistics with oversight provided by Drs. D'Agostino, Reboussin and Liese, thereby providing training opportunities to the next generation of researchers. Aim 3 analyses will be led by Dr. Merchant (mediation/causal inference) and Aim 4 analyses by Dr. Wright (health care costs and expenditures), with guidance from the mPI Drs. Liese and Mendoza, as well as input from Drs. D'Agostino Jr and Reboussin (co-Is).

With respect to training and career development opportunities, the SFS study will not only serve as a training opportunity for pre-doctoral students, but as a career development opportunity for Dr. Sauder, who is a junior faculty member at the University of Colorado and has been one of Dr. Dabelea's mentees. Thus, Drs. Liese and Mendoza will serve as remote mentors to Dr. Sauder, and will set aside dedicated time to consider and discuss various career development opportunities emerging from this effort for Dr. Sauder.

Plan for communication

The investigators have extensive experience undertaking large scale distributed projects spanning multiple clinical sites and institutions. Therefore, this project uses formal project management structure to manage this work. Effective communication is a key component for success. Web-based communication and documentation protocols, along with in-person meetings, conference calls and e-mail will ensure appropriate coordination across sites, committees and individuals. For each regularly scheduled meeting, an agenda and supporting documents will be provided in advance of the meeting, and minutes will be recorded. The PIs will communicate weekly by telephone to discuss protocol implementation, data collection, data analysis, and administrative responsibilities. Since June 2013, this communication plan has been in place and has led to the successful IRB approval and implementation of the data collection procedures for the pilot study at Washington and South Carolina SEARCH sites. Both mPI will share their respective research results with each other and all key personnel and consultants. Each mPI will be responsible for her/his own fiscal and research administration.

Dr. Liese will serve as contact PI and be responsible for submission of progress reports and communications to NIH as well as having overall fiscal responsibility for the study.

Publications

Authorship will be based on the relative scientific contributions of the mPI and key personnel/co-investigators that are part of this application, with inclusion or acknowledgement of SEARCH investigators' contributions as appropriate as per SEARCH 4 publication and presentation policy. Graduate students or post-doctoral fellows working on the project will be given opportunity to contribute to scientific presentations and publications; each trainee will have an individual development plan, which will be described in the progress report.

Conflict Resolution

If a potential conflict develops, the mPI shall meet and attempt to resolve the dispute by reaching consensus. If they fail to resolve the dispute, the disagreement shall be referred to an arbitration committee consisting of one impartial senior executive from each mPI's institution and a third impartial senior executive mutually agreed upon by both mPI. No members of the arbitration committee will be directly involved in the research grant or disagreement.

Change in PI Location

If a mPI moves to a new institution, attempts will be made to transfer the relevant portion of the grant to the new institution. In the event that a mPI cannot carry out his/her duties, a new mPI will be recruited as a replacement at one of the participating institutions.

References

1. American Diabetes Association, "Standards of Medical Care in Diabetes--2016," *Diabetes Care* (2016).

2. American Diabetes Association, "Nutritional Recommendations and Interventions for Diabetes: A Position Statement of the American Diabetes Association," *Diabetes Care* 31, no. Suppl 1 (2008): S61-S78.

3. D. B. Petitti et al., "Glycemic Control in Youth With Diabetes: the SEARCH for Diabetes in Youth Study," *Journal of Pediatrics* 155, no. 5 (2009): 668-672.

4. C. E. de Beaufort et al., "Continuing Stability of Center Differences in Pediatric Diabetes Care: Do Advances in Diabetes Treatment Improve Outcome? The Hvidoere Study Group on Childhood Diabetes," *Diabetes Care* 30, no. 9 (2007): 2245-2250.

5. L. Hanberger et al., "A1C in Children and Adolescents With Diabetes in Relation to Certain Clinical Parameters: the Swedish Childhood Diabetes Registry SWEDIABKIDS," *Diabetes Care* 31, no. 5 (2008): 927-929.

6. Gary Bickel et al., "Guide to Measuring Household Food Security," *Alexandria.Department of Agriculture Food and Nutrition Service* (2000).

7. Meredith Kursmark and Michael Weitzman, "Recent Findings Concerning Childhood Food Insecurity," *Current Opinion in Clinical Nutrition & Metabolic Care* 12, no. 3 (2009): 310-316.

8. H. K. Seligman et al., "Food Insecurity and Glycemic Control Among Low-Income Patients With Type 2 Diabetes," *Diabetes Care* 35, no. 2 (2012): 233-238.

9. Seth A. Berkowitz et al., "Food Insecurity and Metabolic Control Among US Adults With Diabetes," *Diabetes Care* 36, no. 10 (2013): 3093-3099.

10. E. Gucciardi et al., "Exploration of the Relationship Between Household Food Insecurity and Diabetes in Canada," *Diabetes Care* 32, no. 12 (2009): 2218-2224.

11. H. A. Bawadi et al., "Food Insecurity Is Related to Glycemic Control Deterioration in Patients With Type 2 Diabetes," *Clin.Nutr* 31, no. 2 (2012): 250-254.

12. Stacey Marjerrison et al., "Prevalance and Associations of Food Insecurity in Children With Diabetes Mellitus," *The Journal of pediatrics* 158, no. 4 (2011): 607-611.

13. Valerie Tarasuk et al., "Health Care Costs Associated With Household Food Insecurity in Ontario, Canada," *The FASEB Journal* 29, no. 1 Supplement (2015): 261-263.

14. Coleman-Jensen, Alisha, Rabbitt, Matthew P., Gregory, Christian A, and Singh, Anita. Household food security in the United States in 2015. 2016. 9-20-2016.
Ref Type: Report

15. Kelly Stamper Balistreri, "A Decade of Change: Measuring the Extent, Depth and Severity of Food Insecurity," *Journal of Family and Economic Issues* 37, no. 3 (2016): 373-382.

16. Rebekah J. Walker, Joni Strom Williams, and Leonard E. Egede, "Influence of Race, Ethnicity and Social Determinants of Health on Diabetes Outcomes," *The American journal of the medical sciences* 351, no. 4 (2016): 366-373.

17. Dana Dabelea et al., "Etiological Approach to Characterization of Diabetes Type The SEARCH for Diabetes in Youth Study," *Diabetes Care* 34, no. 7 (2011): 1628-1633.

18. Dawn E. Alley et al., "Accountable Health Communities: Addressing Social Needs Through Medicare and Medicaid," *N Engl J Med* 374, no. 1 (2016): 8-11.

19. Benjamin A. Gitterman et al., "Promoting Food Security for All Children," *Pediatrics.* 136, no. 5 (2015): e1431-e1438.

20. World Health Organization and Regional Committee for the Eastern Mediterranean, "Social Determinants of Health," 2016. Available from http://www.slideshare.net/WHO-EMRO/social-determinants-of-health-53566114.

21. Healthy People 2020, "Social determinants of health," 2016. Available from https://www.healthypeople.gov/2020/topics-objectives/topic/social-determinants-of-health.

22. Daniel B. Rubin, Simone R. Singh, and Gary J. Young, "Tax-Exempt Hospitals and Community Benefit: New Directions in Policy and Practice," *Annu.Rev.Public Health* 36 (2015): 545-557.

23. Arvin Garg, Brian Jack, and Barry Zuckerman, "Addressing the Social Determinants of Health Within the Patient-Centered Medical Home: Lessons From Pediatrics," *JAMA* 309, no. 19 (2013): 2001-2002.

24. Arvin Garg et al., "Influence of Maternal Depression on Household Food Insecurity for Low-Income Families," *Academic Pediatrics* 15, no. 3 (2015): 305-310.

25. Dianne Hasselman, "Super-Utilizer Summit: Common Themes From Innovative Complex Care Management Programs," *Center for Health Care Strategies, Inc.Hamilton, NJ* (2013).

26. Holly KM Henry and Edward L. Schor, "Supporting Self-Management of Chronic Health Problems," *Pediatrics.* 135, no. 5 (2015): 789-792.

27. Kelly J. Kelleher et al., "Cost Saving and Quality of Care in a Pediatric Accountable Care Organization," *Pediatrics.* 135, no. 3 (2015): e582-e589.

28. Stephen Berman, "Can Accountable Care Organizations "Disrupt" Our Fragmented Child Health System?," *Pediatrics.* 135, no. 3 (2015): e705-e706.

29. Hoangmai H. Pham, Melissa Cohen, and Patrick H. Conway, "The Pioneer Accountable Care Organization Model: Improving Quality and Lowering Costs," *JAMA* 312, no. 16 (2014): 1635-1636.

30. David J. Nyweide et al., "Association of Pioneer Accountable Care Organizations Vs Traditional Medicare Fee for Service With Spending, Utilization, and Patient Experience," *JAMA* 313, no. 21 (2015): 2152-2161.

31. Winston F. Wong, Thomas A. LaVeist, and Joshua M. Sharfstein, "Achieving Health Equity by Design," *JAMA* 313, no. 14 (2015): 1417-1418.

32. Alastair Baker, "Crossing the Quality Chasm: A New Health System for the 21st Century," *BMJ* 323, no. 7322 (2001): 1192.

33. Dana Dabelea et al., "Prevalence of Type 1 and Type 2 Diabetes Among Children and Adolescents From 2001 to 2009," *JAMA* 311, no. 17 (2014): 1778-1786.

34. Elizabeth J. Mayer-Davis et al., "Trends in the Incidence of Type 1 and Type 2 Diabetes Diagnosed in Youth, 2002-2012: The SEARCH for Diabetes in Youth Study," *N.Engl.J.Med.* (2016): In review.

35. Jean M. Lawrence et al., "Trends in Incidence of Type 1 Diabetes Among Non-Hispanic White Youth in the United States, 2002-2009," *Diabetes.* (2014): DB_131891.

36. A. D. Liese et al., "Neighborhood Level Risk Factors for Type 1 Diabetes in Youth: the SEARCH Case-Control Study," *Int.J Health Geogr* 11 (2012): 1.

37. E. J. Mayer-Davis et al., "The Many Faces of Diabetes in American Youth: Type 1 and Type 2 Diabetes in Five Race and Ethnic Populations: the SEARCH for Diabetes in Youth Study," *Diabetes Care* 32 Suppl 2 (2009): S99-101.

38. J. M. Lawrence et al., "Diabetes in Hispanic American Youth: Prevalence, Incidence, Demographics, and Clinical Characteristics: the SEARCH for Diabetes in Youth Study," *Diabetes Care* 32 Suppl 2 (2009): S123-S132.

39. R. A. Bell et al., "Diabetes in Non-Hispanic White Youth: Prevalence, Incidence, and Clinical Characteristics: the SEARCH for Diabetes in Youth Study," *Diabetes Care* 32 Suppl 2 (2009): S102-S111.

40. Coleman-Jensen, A., Gregory, C, and Singh, A. Household Food Security in the United States in 2013. ERR-173, i-33. 2014. Economic Research Service,U.S.Department of Agriculture. Economic Research Report. 11-8-2014.
Ref Type: Report

41. Heather A. Eicher-Miller et al., "Food Insecurity Is Associated With Diet and Bone Mass Disparities in Early Adolescent Males but Not Females in the United States," *The Journal of nutrition* 141, no. 9 (2011): 1738-1745.

42. K. Alaimo, C. M. Olson, and E. A. Frongillo, Jr., "Low Family Income and Food Insufficiency in Relation to Overweight in US Children: Is There a Paradox?," *Arch Pediatr Adolesc Med* 155, no. 10 (2001): 1161-1167.

43. Neil W. Boris et al., "Depressive Symptoms in Youth Heads of Household in Rwanda: Correlates and Implications for Intervention," *Archives of pediatrics & adolescent medicine* 162, no. 9 (2008): 836-843.

44. M. A. Broughton et al., "Predictors and Outcomes of Household Food Insecurity Among Inner City Families With Preschool Children in Vancouver," *Can J Public Health* 97 (2006): 214-216.

45. Patrick Casey et al., "Maternal Depression, Changing Public Assistance, Food Security, and Child Health Status," *Pediatrics.* 113, no. 2 (2004): 298-304.

46. John T. Cook et al., "Are Food Insecurity's Health Impacts Underestimated in the US Population? Marginal Food Security Also Predicts Adverse Health Outcomes in Young US Children and Mothers," *Advances in Nutrition: An International Review Journal* 4, no. 1 (2013): 51-61.

47. J. T. Cook and D. A. Frank, "Food Security, Poverty, and Human Development in the United States," *Annals of the New York Academy of Sciences* 2007/10/24 (2008): 193-209.

48. J. T. Cook et al., "Food Insecurity Is Associated With Adverse Health Outcomes Among Human Infants and Toddlers," *J Nutr* 134, no. 6 (2004): 1432-1438.

49. R. E. Kleinman et al., "Hunger in Children in the United States: Potential Behavioral and Emotional Correlates," *Pediatrics* 1998/01/07 (1998): E3.

50. J. M. Murphy et al., "Relationship Between Hunger and Psychosocial Functioning in Low-Income American Children," *Journal of the American Academy of Child and Adolescent Psychiatry* 1998/02/25 (1998): 163-170.

51. R Rose-Jacobs et al., "Household Food Insecurity: Associations With at-Risk Infant and Toddler Development," *Pediatrics* 121, no. 1 (2008): 65-72.

52. R. C. Whitaker, S. M. Phillips, and S. M. Orzol, "Food Insecurity and the Risks of Depression and Anxiety in Mothers and Behavior Problems in Their Preschool-Aged Children," *Pediatrics* 118, no. 3 (2006): e859-e868.

53. M. Zaslow et al., "Food Security During Infancy: Implications for Attachment and Mental Proficiency in Toddlerhood," *Maternal and child health journal* 2008/03/05 (2009): 66-80.

54. D. F. Jyoti, E. A. Frongillo, and S. J. Jones, "Food Insecurity Affects School Children's Academic Performance, Weight Gain, and Social Skills," *The Journal of nutrition* 2005/12/01 (2005): 2831-2839.

55. K. Alaimo et al., "Food Insufficiency, Family Income, and Health in US Preschool and School-Aged Children," *Am.J Public Health* 91, no. 5 (2001): 781-786.

56. Katherine Alaimo, Christine M. Olson, and Edward A. Frongillo, "Family Food Insufficiency, but Not Low Family Income, Is Positively Associated With Dysthymia and Suicide Symptoms in Adolescents," *The Journal of nutrition* 132, no. 4 (2002): 719-725.

57. Hilary K. Seligman et al., "Food Insecurity and Hypoglycemia Among Safety Net Patients With Diabetes," *Arch.Intern.Med.* 171, no. 13 (2011): 1204-1206.

58. John Billimek and Dara H. Sorkin, "Food Insecurity, Processes of Care, and Self-Reported Medication Underuse in Patients With Type 2 Diabetes: Results From the California Health Interview Survey," *Health Serv.Res.* 47, no. 6 (2012): 2159-2168.

59. Julie Silverman et al., "The Relationship Between Food Insecurity and Depression, Diabetes Distress and Medication Adherence Among Low-Income Patients With Poorly-Controlled Diabetes," *J.Gen.Intern.Med.* (2015): 1-5.

60. Court Lyles et al., "Food Insecurity in Relation to Changes in Hemoglobin A1c, Self-Efficacy, and Fruit/Vegetable Intake During a Diabetes Educational Intervention," *Diabetes Care* 36, no. 6 (2013): 1448-1453.

61. Seth A. Berkowitz, Hilary K. Seligman, and Niteesh K. Choudhry, "Treat or Eat: Food Insecurity, Cost-Related Medication Underuse, and Unmet Needs," *The American journal of medicine* 127, no. 4 (2014): 303-310.

62. Matthew M. Ippolito et al., "Food Insecurity and Diabetes Self-Management Among Food Pantry Clients," *Public Health Nutrition* (2016): In press.

63. Chadwick K. Knight et al., "Household Food Insecurity and Medication "Scrimping" Among US Adults With Diabetes," *Preventive Medicine* 83 (2016): 41-45.

64. Sheri D. Weiser et al., "Conceptual Framework for Understanding the Bidirectional Links Between Food Insecurity and HIV/AIDS," *The American Journal of Clinical Nutrition* 94, no. 6 (2011): 1729S-1739S.

65. Sheri D. Weiser et al., "Food insecurity and health: A conceptual framework," in *Food Insecurity and Public Health*, ed. L Ivers. (Boca Raton: CDC Press, 2015), 23-50.

66. Daria R. Homenko et al., "Food Insecurity and Food Choices in Rural Older Adults With Diabetes Receiving Nutrition Education Via Telemedicine," *Journal of Nutrition Education and Behavior* 42, no. 6 (2010): 404-409.

67. A. Kendall, C. M. Olson, and E. A. Frongillo, Jr., "Relationship of Hunger and Food Insecurity to Food Availability and Consumption," *J Am.Diet.Assoc.* 96, no. 10 (1996): 1019-1024.

68. Jayanta Bhattacharya, Janet Currie, and Steven Haider, "Poverty, Food Insecurity, and Nutritional Outcomes in Children and Adults," *J.Health Econ.* 23, no. 4 (2004): 839-862.

69. Catherine M. Champagne et al., "Poverty and Food Intake in Rural America: Diet Quality Is Lower in Food Insecure Adults in the Mississippi Delta," *J.Am.Diet.Assoc.* 107, no. 11 (2007): 1886-1894.

70. Catherine Huet, Renata Rosol, and Grace M. Egeland, "The Prevalence of Food Insecurity Is High and the Diet Quality Poor in Inuit Communities," *The Journal of nutrition* 142, no. 3 (2012): 541-547.

71. H. K. Seligman et al., "Food Insecurity Is Associated With Hypoglycemia and Poor Diabetes Self-Management in a Low-Income Sample With Diabetes," *J Health Care Poor Underserved* 21, no. 4 (2010): 1227-1233.

72. Seth A. Berkowitz et al., "Material Need Insecurities, Control of Diabetes Mellitus, and Use of Health Care Resources: Results of the Measuring Economic Insecurity in Diabetes Study," *JAMA internal medicine* 175, no. 2 (2015): 257-265.

73. Lily S. Barnard et al., "Material Need Support Interventions for Diabetes Prevention and Control: a Systematic Review," *Current diabetes reports* 15, no. 2 (2015): 1-8.

74. Quyen G. To et al., "Household Food Insecurity Is Associated With Less Physical Activity Among Children and Adults in the US Population," *The Journal of nutrition* 144, no. 11 (2014): 1797-1802.

75. Maya Vijayaraghavan et al., "The Association Between Housing Instability, Food Insecurity, and Diabetes Self-Efficacy in Low-Income Adults," *J.Health Care Poor Underserved* 22, no. 4 (2011): 1279-1291.

76. Tyler VanderWeele, "Multiple Mediators: A weighting approach to multiple mediators," in *Explanation in Causal Inference: Methods for Mediation and Interaction* (New York: Oxford University Press, 2015), 123-125.

77. Robert R. Althoff, Merelise Ametti, and Farryl Bertmann, "The Role of Food Insecurity in Developmental Psychopathology," *Preventive Medicine* (2016).

78. Ashley L. Munger, Sandra L. Hofferth, and Stephanie K. Grutzmacher, "The Role of the Supplemental Nutrition Assistance Program in the Relationship Between Food Insecurity and Probability of Maternal Depression," *Journal of Hunger & Environmental Nutrition* 11, no. 2 (2016): 147-161.

79. Arvin Garg et al., "Influence of Maternal Depression on Household Food Insecurity for Low-Income Families," *Academic Pediatrics* 15, no. 3 (2015): 305-310.

80. Center on Budget and Policy Priorities, *Policy Basics: Introduction to SNAP*: Washington DC, 2014).

81. C. L. Connell et al., "Children's Experiences of Food Insecurity Can Assist in Understanding Its Effect on Their Well-Being," *J Nutr* 135, no. 7 (2005): 1683-1690.

82. United States Department of Agriculture, "Estimates of State Supplemental Nutrition Assistance Program participation rates in 2013," 2016. Available from http://www.fns.usda.gov/sites/default/files/ops/Reaching2013.pdf.

83. United States Department of Agriculture, "Trends in Supplemental Nutrition Assistance Program participation rates: Fiscal year 2010 to fiscal year 2014 (summary)," 2016. Available from http://www.fns.usda.gov/sites/default/files/ops/Trends2010-2014-Summary.pdf.

84. Fox, M. K. Effects of Food Assistance and Nutrition Programs on Nutrition and Health, Volume 4 Executive Summary of the Literature. Economic Research Service, United States Department of Agriculture, Food Assistance & Nutrition Research Program 19-4. 2004. Economic Research Services, United States Department of Agriculture.
Ref Type: Report

85. Enza Gucciardi et al., "The Intersection Between Food Insecurity and Diabetes: a Review," *Current nutrition reports* 3, no. 4 (2014): 324-332.

86. Rebekah J. Walker, Brittany L. Smalls, and Leonard E. Egede, "Social Determinants of Health in Adults With Type 2 Diabetes-Contribution of Mutable and Immutable Factors," *Diabetes Res.Clin.Pract.* 110, no. 2 (2015): 193-201.

87. W. J. Heerman et al., "Food Insecurity Is Associated With Diabetes Self-Care Behaviours and Glycaemic Control," *Diabet.Med.* (2015).

88. Ivy Shiue, "People With Diabetes, Respiratory, Liver or Mental Disorders, Higher Urinary Antimony, Bisphenol A, or Pesticides Had Higher Food Insecurity: USA NHANES, 2005-2006," *Environmental Science and Pollution Research* 23, no. 1 (2016): 198-205.

89. Meng Ding et al., "Undiagnosed Prediabetes Among Food Insecure Adults," *American Journal of Health Behavior* 38, no. 2 (2014): 225-233.

90. Junxiu Liu et al., "Gender Differences in the Association Between Food Insecurity and Insulin Resistance Among US Adults: National Health and Nutrition Examination Survey, 2005-2010," *Annals of Epidemiology* 25, no. 9 (2015): 643-648.

91. Majid Hasan-Ghomi et al., "Relationship of Food Security With Type 2 Diabetes and Its Risk Factors in Tehranian Adults," *International journal of preventive medicine* 6 (2015): 98.

92. Sabrina Strings et al., "Race and Sex Differences in the Association Between Food Insecurity and Type 2 Diabetes," *Ethnicity & disease* 26, no. 3 (2016): 427-434.

93. Yixing Sun, "The Association of Food Security Status With Diabetes in American Adolescents," (2011).

94. Rafael Perez-Escamilla et al., "Household Food Insecurity, Diabetes and Hypertension Among Mexican Adults: Results From Ensanut 2012," *salud pública de méxico* 56 (2014): s62-s70.

95. TODAY Study Group, "Treatment Options for Type 2 Diabetes in Adolescents and Youth: a Study of the Comparative Efficacy of Metformin Alone or in Combination With Rosiglitazone or Lifestyle Intervention in Adolescents With Type 2 Diabetes," *Pediatric diabetes* 8, no. 2 (2007): 74.

96. Kenneth C. Copeland et al., "Characteristics of Adolescents and Youth With Recent-Onset Type 2 Diabetes: the TODAY Cohort at Baseline," *The Journal of Clinical Endocrinology & Metabolism* 96, no. 1 (2011): 159-167.

97. Carolyn E. Ievers-Landis et al., "Longitudinal Correlates of Health Risk Behaviors in Children and Adolescents With Type 2 Diabetes," *The Journal of pediatrics* 166, no. 5 (2015): 1258-1264.

98. Rachelle Gandica and Phil Zeitler, "Update on Youth-Onset Type 2 Diabetes: Lessons Learned From the Treatment Options for Type 2 Diabetes in Adolescents and Youth Clinical Trial," *Adv.Pediatr.* (2016).

99. Fram MS et al., "Children Are Aware of Food Insecurity and Take Responsiblity for Managing Resources," *J Nutr* 141, no. 6 (2011): 1114-1119.

100. Maryah Stella Fram et al., "Development and Validation of a Child Report Assessment of Child Food Insecurity and Comparison to Parent Report Assessment," *Journal of Hunger & Environmental Nutrition* 8, no. 2 (2013): 128-145.

101. Mary Carol Burkhardt et al., "Enhancing Accurate Identification of Food Insecurity Using Quality-Improvement Techniques," *Pediatrics.* 129, no. 2 (2012): e504-e510.

102. Jennifer Bernal et al., "Food Insecurity in Children but Not in Their Mothers Is Associated With Altered Activities, School Absenteeism, and Stunting," *The Journal of nutrition* 144, no. 10 (2014): 1619-1626.

103. Hilary K. Seligman, Barbara A. Laraia, and Margot B. Kushel, "Food Insecurity Is Associated With Chronic Disease Among Low-Income NHANES Participants," *The Journal of nutrition* 140, no. 2 (2010): 304-310.

104. Michael P. Burke et al., "Severity of Household Food Insecurity and Lifetime Racial Discrimination Among African-American Households in South Carolina," *Ethnicity & Health* (2016): 1-17.

105. Irma Arteaga, Stephanie Potochnick, and Sarah Parsons, "Decomposing the Household Food Insecurity Gap for Children of US-Born and Foreign-Born Hispanics: Evidence From 1998 to 2011," *Journal of Immigrant and Minority Health* (2017): 1-9.

106. Justin T. Denney et al., "Social Cohesion and Food Insecurity: Insights From the Geographic Research on Wellbeing (GROW) Study," *Maternal and child health journal* (2016): 1-8.

107. SL Pruitt et al., "Who Is Food Insecure? Inplications for Targeted Recruitment and Outreach, National Health and Nutrition Examination Survey, 2005-2010," *Preventing Chronic Disease* 13 (2016): e143.

108. A. D. Liese et al., "Participation in Pediatric Epidemiologic Research: the SEARCH for Diabetes in Youth Study Experience," *Contemp.Clin Trials* 29, no. 6 (2008): 829-836.

109. Steven J. Carlson, Margaret S. Andrews, and Gary W. Bickel, "Measuring Food Insecurity and Hunger in the United States: Development of a National Benchmark Measure and Prevalence Estimates," *The Journal of nutrition* 129, no. 2 (1999): 510S-516S.

110. E. A. Frongillo, Jr., "Validation of Measures of Food Insecurity and Hunger," *J Nutr* 129, no. 2S Suppl (1999): 506S-509S.

111. Jane L. Chiang et al., "Type 1 Diabetes Through the Life Span: A Position Statement of the American Diabetes Association," *Diabetes Care* 37, no. 7 (2014): 2034-2054.

112. Marian J. Rewers et al., "Assessment and Monitoring of Glycemic Control in Children and Adolescents With Diabetes," *Pediatric diabetes* 15 (2014): 102-114.

113. SEARCH Study Group, "SEARCH for Diabetes in Youth: a Multicenter Study of the Prevalence, Incidence and Classification of Diabetes Mellitus in Youth," *Control Clin Trials* 25, no. 5 (2004): 458-471.

114. Allison Larg and John R. Moss, "Cost-of-Illness Studies," *Pharmacoeconomics* 29, no. 8 (2011): 653-671.

115. Anita Patel et al., "A Comparison of Two Methods of Collecting Economic Data in Primary Care," *Fam.Pract.* 22, no. 3 (2005): 323-327.

116. Davene R. Wright et al., "The Cost of a Primary Care-Based Childhood Obesity Prevention Intervention," *BMC health services research* 14, no. 1 (2014): 44.

117. "Dietary Screener Questionnaire in the National Health Interview Survey Cancer Control Supplement 2015,", National Cancer Institute Division of Cancer Control and Populatin Sciences (2018; <[28] Date).

118. Lenore Sawyer Radloff, "The CES-D Scale a Self-Report Depression Scale for Research in the General Population," *Applied psychological measurement* 1, no. 3 (1977): 385-401.

119. Jerry L. Rushton, Michelle Forcier, and Robin M. Schectman, "Epidemiology of Depressive Symptoms in the National Longitudinal Study of Adolescent Health," *Journal of the American Academy of Child & Adolescent Psychiatry* 41, no. 2 (2002): 199-205.

120. Robert L. Spitzer et al., "A Brief Measure for Assessing Generalized Anxiety Disorder: the GAD-7," *Arch.Intern.Med.* 166, no. 10 (2006): 1092-1097.

121. S. Cohen, T. Kamarck, and R. Mermelstein, "A Global Measure of Perceived Stress Journal of Health and Social Behavior, 24," *Journal of Health and Social Bahavior* 24, no. 4 (1983): 385-396.

122. Alison L. Marshall et al., "Measuring Total and Domain-Specific Sitting: a Study of Reliability and Validity," *Med.Sci.Sports Exerc.* 42, no. 6 (2010): 1094-1102.

123. C. L. Craig et al., "International Physical Activity Questionnaire: 12-Country Reliability and Validity," *Medicine & Science in Sports & Exercise* 35, no. 8 (2003): 1381-1395.

124. S. Cohen, *The meaning and measurement of social support* (New York, NY: Hemisphere Publishing Corporation, 1992).

125. Gregory D. Zimet et al., "The Multidimensional Scale of Perceived Social Support," *J.Pers.Assess.* 52, no. 1 (1988): 30-41.

126. Catherine Pihoker et al., "Insulin Regimens and Clinical Outcomes in a Type 1 Diabetes Cohort," *Diabetes Care* 36, no. 1 (2013): 27-33.

127. US Department of Housing and Urban Development, "Homeless Emergency Assistance and Rapid Transition to Housing: Rural Housing Stability Assistance Program and Revisions to the Definition of "Chronically Homeless", 24 CFR 579," *Fed.Regist.* 78, no. 59 (2013): 18726-18761.

128. A. D. Liese et al., "Recruitment Strategies and Participation in a Study of Childhood Hunger," *Journal of Hunger & Environmental Nutrition* (2014): 1-18.

129. S. J. Jones et al., "Stressful Life Events, Economic and Social Capabilities, and Competing Demands: Associations With Food Insecurity and Child Hunger," *American Journal of Epidemiology* (2014): In review.

130. MM Huizinga et al., "Development and Validation of the Diabetes Numeracy Test (DNT)," *BMC health services research* 8, no. 1 (2008): 96.

131. D. R. Williams et al., "Racial Differences in Physical and Mental Health: Socioeconomic Status, Stress, and Discrimination," *Journal of Health Psychology* 2, no. 3 (1997): 335-351.

132. Center for Disease Control and Prevention, "2003-2004 Current National Health and Nutrtion Examination Survery (NHANES 2003-2004)," 2016. Available from http://wwwn.cdc.gov/nchs/nhanes/search/nhanes03_04.aspx.

133. Sanjay Basu et al., "Ending SNAP Subsidies for Sugar-Sweetened Beverages Could Reduce Obesity and Type 2 Diabetes," *Health Aff.(Millwood).* 33, no. 6 (2014): 1032-1039.

134. Marlana J. Kohn et al., "Food Insecurity, Food Assistance and Weight Status in US Youth: New Evidence From NHANES 2007GÇô08," *Pediatric obesity* 9, no. 2 (2014): 155-166.

135. Cindy W. Leung et al., "Associations of Food Stamp Participation With Dietary Quality and Obesity in Children," *Pediatrics.* 131, no. 3 (2013): 463-472.

136. U. Pramming S. & Thorsteinsson B. Pedersen-Bjergaard, "Recall of Severe Hypoglycaemia and Self-Estimated State of Awareness in Type 1 Diabetes," *Diabetes/Metabolism Research and Reviews* 19, no. 3 (2003): 232-240.

137. R. Aronson et al., "The Canadian Hypoglycemia Assessment Tool Program: Insights Into Rates and Implications of Hypoglycemia From an Observational Study," *Canadian journal of diabetes* 42, no. 1 (2018): 11-17.

138. Davene R. Wright et al., "Modeling the Cost Effectiveness of Child Care Policy Changes in the US," *Am.J.Prev.Med.* 49, no. 1 (2015): 135-147.

139. DR Wright et al., "The Costs and Cost-Effectiveness of Collaborative Care for Adolescents With Depression in Primary Care Settings," *JAMA Pediatrics* (2016): In print.

140. Agency for Healthcare Research and Quality, "Medical Expenditure Panel Survey," 2016. Available from https://meps.ahrq.gov/mepsweb/.

141. United States Department of Labor, "Bureau of Labor Statistics: Consumer price index," 2016. Available from http://www.bls.gov/cpi/.

142. United States Department of Labor, "Bureau of labor statistics: May 2015 national occupational employment and wage estimaets United States," 2016. Available from http://www.bls.gov/oes/current/oes_nat.htm#00-0000.

143. Scott D. Grosse, Kurt V. Krueger, and Mercy Mvundura, "Economic Productivity by Age and Sex: 2007 Estimates for the United States," *Med.Care* (2009): S94-S103.

144. Merjaneh, Lina, Pihoker, Catherine, Fino, Nora, Divers, Jasmin, Klingensmith, Georgeanna, Shrestha, Sundar, Savdah, Sharon, Mayer-Davis, Elizabith J., Dabelea, Dana, Powell, Jeffrey, Lawrence, Jean M., Dolan, Lawrence, Linder, Barbara, Wright, Davene, and For the SEARCH for Diabetes in Youth Study Group. Out of pocket medical expenses for children and young adults with diabetes: The SEARCH for diabetes in Youth Cohort Study. ADA 76th Scientific Sessions, New Orleans . 6-11-2016.
Ref Type: Abstract

145. World Health Organization, "Validity and comparability of out-of-pocket health expenditure from household surveys: a review of the literature and current survey instruments," 2016. Available from http://www.who.int/health_financing/documents/dp_e_11_01-oop_errors.pdf.

146. J. M. Lawrence et al., "Prevalence and Correlates of Depressed Mood Among Youth With Diabetes: the SEARCH for Diabetes in Youth Study," *Pediatrics* 117, no. 4 (2006): 1348-1358.

147. Eaton, W. W., Smith, C., Ybarra, M., Muntaner, C., and Tien, A. Center for Epidemiologic Studies Depression Scale: review and revision (CESD and CESD-R). 2004.
Ref Type: Report

148. Bernd Lowe et al., "Validation and Standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the General Population," *Med.Care* 46, no. 3 (2008): 266-274.

149. Emma Robinson et al., "Internet Treatment for Generalized Anxiety Disorder: a Randomized Controlled Trial Comparing Clinician Vs. Technician Assistance," *PLoS One* 5, no. 6 (2010): e10942.

150. Janine Clarke et al., "A Mobile Phone and Web-Based Intervention for Improving Mental Well-Being in Young People With Type 1 Diabetes: Design of a Randomized Controlled Trial," *JMIR research protocols* 4, no. 2 (2015).

151. Lisa Robins et al., "Internet-Delivered Cognitive Behaviour Therapy for Depression in People With Diabetes: Study Protocol for a Randomised Controlled Trial," *BMJ open diabetes research & care* 3, no. 1 (2015): e000144.

152. S. J. Jones et al., "Stressful Life Events, Economic and Social Capabilities, and Competing Demands: Associations With Food Insecurity and Child Hunger," *American Journal of Epidemiology* (2014): In review.

153. S. J. Jones et al., "Stressful Life Events, Economic and Social Capabilities, and Competing Demands: Associations With Food Insecurity and Child Hunger," *American Journal of Epidemiology* (2014): In review.

154. S. J. Jones et al., "Stressful Life Events, Economic and Social Capabilities, and Competing Demands: Associations With Food Insecurity and Child Hunger," *American Journal of Epidemiology* (2014): In review.

155. National Cancer Institute Epidemiology and Genomics Research Program, "Dietary Screener Questionnaire in the National Health Interview Survey Cancer Control Supplement 2015," 2018. Available from https://epi.grants.cancer.gov/nhis/2015-screener/.

156. C. L. Craig et al., "International Physical Activity Questionnaire: 12-Country Reliability and Validity," *Medicine & Science in Sports & Exercise* 35, no. 8 (2003): 1381-1395.

157. C. L. Craig et al., "International Physical Activity Questionnaire: 12-Country Reliability and Validity," *Medicine & Science in Sports & Exercise* 35, no. 8 (2003): 1381-1395.

158. C. L. Craig et al., "International Physical Activity Questionnaire: 12-Country Reliability and Validity," *Medicine & Science in Sports & Exercise* 35, no. 8 (2003): 1381-1395.

159. C. L. Craig et al., "International Physical Activity Questionnaire: 12-Country Reliability and Validity," *Medicine & Science in Sports & Exercise* 35, no. 8 (2003): 1381-1395.

160. Paul H. Lee et al., "Validity of the International Physical Activity Questionnaire Short Form (IPAQ-SF): A Systematic Review," *International Journal of Behavioral Nutrition and Physical Activity* 8, no. 1 (2011): 1.

161. SEARCH Study Group, "SEARCH for Diabetes in Youth: a Multicenter Study of the Prevalence, Incidence and Classification of Diabetes Mellitus in Youth," *Controlled Clinical Trials* 25, no. 5 (2004): 458-471.

162. S. J. Jones et al., "Child Hunger From a Family Resilience Perspective," *Journal of Hunger & Environmental Nutrition* (2018): 1-22.

163. Cavanaugh K et al., "Association of Numeracy and Diabetes Control," *Annals of Internal Medicine* 148, no. 10 (2008): 737-746.

164. "
THE DNT 5: A Shortened Version of the Diabetes Numeracy Test (DNT)," Available from www.huskyhealthct.org/providers/provider_postings/diabetes/DNT5.pdf.

165. Candace D McNaughton et al., "Validation of a Short, 3-Item Version of the Subjective Numeracy Scale," *Medical Decision Making* 35, no. 8 (2015): 932-936.

166. Administration for Children and Families Office of Family Assistance DHHS. Enhancing Family Stability: A Guide for Addessing Housing Status and Coordinating with Local Homelessness Programs for TANF Agencies. 2016.
Ref Type: Report

167. National Health Care for the Homeless Council, "Frequently Asked Questions,".

168. National Alliance to End Homelessness, "Changes in the HUD Definition of "Homeless"," Available from https://endhomelessness.org/resource/changes-in-the-hud-definition-of-homeless/.

169. J Hintze, *PASS 13* (Kaysville, Utah: NCSS, LLC, 2014).

170. Ralph B. D'Agostino, "Tutorial in Biostatistics: Propensity Score Methods for Bias Reduction in the Comparison of a Treatment to a Non-Randomized Control Group," *Stat Med* 17, no. 19 (1998): 2265-2281.

171. Alison Snow Jones et al., "Assessing the Effect of Batterer Program Completion on Reassault Using Propensity Scores," *Journal of Interpersonal Violence* 19, no. 9 (2004): 1002-1020.

172. Ellice Lieberman et al., "Association of Epidural Analgesia With Cesarean Delivery in Nulliparas," *Obstetrics & Gynecology* 88, no. 6 (1996): 993-1000.

173. DaLi Feng et al., "Hemostatic State and Atrial Fibrillation (the Framingham Offspring Study)," *The American journal of cardiology* 87, no. 2 (2001): 168-171.

174. Robin L. Kruse et al., "Does Hospitalization Impact Survival After Lower Respiratory Infection in Nursing Home Residents?," *Med.Care* 42, no. 9 (2004): 860-870.

175. Ralph B. D'Agostino, Jr., "Propensity Scores in Cardiovascular Research," *Circulation* 115, no. 17 (2007): 2340-2343.

176. Ronan Roussel et al., "Metformin Use and Mortality Among Patients With Diabetes and Atherothrombosis," *Arch.Intern.Med.* 170, no. 21 (2010): 1892-1899.

177. Lars G. Svensson et al., "Results of Matching Valve and Root Repair to Aortic Valve and Root Pathology," *The Journal of thoracic and cardiovascular surgery* 142, no. 6 (2011): 1491-1498.

178. Raghava S. Velagaleti et al., "Left Ventricular Hypertrophy Patterns and Incidence of Heart Failure With Preserved Versus Reduced Ejection Fraction," *The American journal of cardiology* 113, no. 1 (2014): 117-122.

179. Ralph B. D'Agostino, Jr., "Overview of Missing Data Techniques," *Topics in Biostatistics* (2007): 339-352.

180. Ralph D'Agostino, Jr. et al., "Examining the Impact of Missing Data on Propensity Score Estimation in Determining the Effectiveness of Self-Monitoring of Blood Glucose (SMBG)," *Health Services and Outcomes Research Methodology* 2, no. 3-4 (2001): 291-315.

181. Ralph B. D'Agostino, Jr. and Donald B. Rubin, "Estimating and Using Propensity Scores With Partially Missing Data," *J Am Stat Assoc* 95, no. 451 (2000): 749-759.

182. D. Ruppert, M. P. Wand, and R. J. Carroll, *Semi-parametric Regression*: Cambridge University Press, 2003).

183. US Department of Health and Human Services, "2013 Poverty Guidelines," 2014. Available from https://aspe.hhs.gov/2013-poverty-guidelines.

184. Kristopher J. Preacher, "Advances in Mediation Analysis: A Survey and Synthesis of New Developments," *Annu.Rev.Psychol.* 66 (2015): 825-852.

185. R. M. Baron and D. A. Kenny, "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations," *Journal of Personality and Social Psychology* 51, no. 6 (1986): 1173-1182.

186. Samuel H. Zuvekas and Gary L. Olin, "Validating Household Reports of Health Care Use in the Medical Expenditure Panel Survey," *Health Serv.Res.* 44, no. 5p1 (2009): 1679-1700.

187. Annie Qu and Peter XGÇÉK Song, "Testing Ignorable Missingness in Estimating Equation Approaches for Longitudinal Data," *Biometrika* 89, no. 4 (2002): 841-850.

Additional references:

Jessica Markowitz et al., "Brief Screening Tool for Disordered Eating in Diabetes," *Diabetes Care* 33, no. 3 (2010): 495-500.

Kathryn M. Connor and Jonathan R.T. Davidson, "Development of a New Resilience Scale: The Connor-Davison Resilience Scale (CD-RISC)," *Depression and Anxiety* 18 (2003): 76-82.