SEARCH 4 MOP
Section 20 - Prevalence of Diabetes in Youth in 2017
Table of Contents

20. Prevalence of Diabetes in Youth in 2017

20.1. RATIONALE AND GOAL

The goals of the "2017 Prevalence Study" are 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 prevalence areas/health plan/IHS in 2017 in order to:

a) estimate the population prevalence of diabetes in youth age < 20 years in 2017 by age, sex, race/ethnicity and diabetes type; and

b) compare the prevalence of diabetes in youth age < 20 years in 2001, 2009 and 2017, by age, sex, race/ethnicity and diabetes type.

Given these goals, the methods of case ascertainment, validation, data collection of core variables, and denominator estimation need to be similar to those employed for the 2001 and 2009 Prevalence Studies.

20.2. METHODS: CASE ASCERTAINMENT

This section provides a general overview of the methods that will be used to accomplish the goals. More detailed site-specific information is included in Appendix 1: Site-Specific Case Ascertainment Methods

20.2.1. *Case Identification*

20.2.1.1. Validation Criteria

Physician diagnosis of diabetes (not including gestational diabetes [GDM] only).

20.2.1.2. Eligibility Criteria

The eligibility criteria for prevalent cases of diabetes are as follows:

- diabetes (other than gestational diabetes only) was diagnosed during or prior to 2017;

- age less than 20 years on December 31, 2017. This corresponds to a birth year of 01/01/98 – 12/31/2017 for the prevalence year of 2017. Subjects reaching age 20 years during the prevalence year are not eligible;

- resident of the population defined for prevalent cases at any time during the prevalence year (for geographically-based centers) or member of the participating health plan or IHS region at any time during the prevalence year;

- not active-duty military;

- not living in an institution (defined as such by the Census).

For **geographically-based centers**, residency is defined as anyone who resided in that center's geographical area at any time during the prevalence year. It is expected that residence eligibility will be determined at the time a case is identified. If this is not possible, residence eligibility needs to be validated after case identification.

For the **membership-based center**, membership is defined as anyone who was a member of the health plan at any time during the prevalence year. It is expected that membership eligibility will be determined at the time the case is identified. If this is not possible, membership eligibility needs to be validated after case identification.

Prevalent cases of diabetes that are **active duty military personnel** _are ineligible_ because it is anticipated that access to military medical records will be difficult. It is unlikely that the prevalent diabetes cases will be present among active duty military because diabetes precludes active duty military service.

Prevalent cases of diabetes that are **dependents of military personnel** having access to civilian medical facilities _are eligible_. However, prevalent diabetes cases receiving care only at military facilities will be ascertained only if access to military records is obtained.

Prevalent cases of diabetes that are **college students** _are eligible_ as these youth are counted in the Census as resident at their college address. Additionally, college students identified as cases by the membership-based center cannot be removed from the membership-based administrative records that are the source of denominator data for these centers.

It is expected that some of the 2017 prevalent cases will have previously been identified by SEARCH as prevalent in 2001 or 2009 or incident in 2002-2017. Those identified exclusively as being prevalent in 2017 will include persons that were missed by SEARCH in the previous case ascertainment windows, persons that moved into the catchment area after being diagnosed with diabetes in another location, and persons that joined the health plan after being diagnosed with diabetes.

All 2017 incident cases that reside in the prevalence areas will also be 2017 prevalent cases. Some of the 2001 or 2009 prevalent and 2002-2016 incident cases will not be identified as 2017 prevalent for one or more of the following reasons: a) have reached age 20 before December 31, 2017; b) were not resident of the geographically-defined area _at any time_ in 2017; c) were not a member of the health plan _at any time_ during 2017; d) died before 1/1/2017; e) became active duty military before 1/1/2017; or f) they entered an institution before 1/1/2017.

20.2.2. *Case Finding Approaches*

The approach to case-finding for prevalent cases will vary by center. All of the data sources that were used to identify cases prevalent in 2001 and 2009 and incident in 2002-2017 will be used to identify prevalent 2017 cases if they are still available. These will include both clinical and administrative sources, such as health care provider case reports; hospital discharge records, laboratory test results, prescription dispensed, outpatient clinical records (including pediatric and adult endocrine practices and primary care practices) and encounter codes; established diabetes registries; and state insurance plans. The SEARCH tracking database should not be used as a source for ascertaining cases to be consistent with ascertainment in the prevalence 2001 and 2009 cases.

For geographically based centers, database searches for potential cases will include the previous three years, retrospective to January 1, 2014 (to replicate the 2001 and 2009 studies).

Administrative data, when available, will be used to verify eligibility, including geographic, membership, and age-related criteria. The following information is needed in order to correctly identify all eligible, unique cases of diabetes:

- o *name (or at least partial name, initials, etc.)*
- o *gender/sex*
- o *date of birth*
- o *zip code and/or county of residence in the prevalence year (if possible full address and phone number for contact information purposes)*
- o *health plan membership during 2017*
- o *IHS beneficiary in 2017*

If date of birth and zip code/county of residence are not available, they will have to be collected later when the initial participant survey (IPS) is administered or when the core form is completed.

In addition, the following "core" information is ideally needed to describe cases and to facilitate future contact with the participant:

- o *diabetes type as determined by provider in 2017 (e.g., from clinical databases, the notes in the electronic health records, or provider case reports)*
- o *race/ethnicity*
- o *date of diagnosis*
- o *name, address, fax/phone/email of current primary care and/or diabetes care provider*

The center-specific protocols for case ascertainment (Appendix 1) provide more detail on the case finding approaches for each SEARCH center.

Table 1 summarizes the data sources that will be used at each center to identify prevalent cases.

Table 1. Data Sources Used to Identify Prevalent Cases by Center

Center	Hospital Discharge	Laboratory	Prescription	Ambulatory Billing or Encounter Codes or Dx Codes	Pediatric Endocrinology Case List	Other
Ohio	X			X	X	X
Colorado	X			X	X	X
Washington	X			X	X	X
Carolinas	X			X	X	X
California	X	X	X	X	X	X

20.2.3. *Identification and Elimination of Possible Duplicate Cases*

Cases will be identified based on data sources available at each individual center. Once cases are identified, it is necessary to identify and eliminate duplicate cases. This involves establishing a method to match records to identify duplicates at each center.

The California center will use unique identifying information including medical record number, name, and date of birth to identify and remove duplicates records. The Ohio center has a registration system for cases that will permit duplicates to be identified as they are entered into the diabetes database based on name, date of birth, and other identifying information about the child and parents. The other three centers (Colorado, Washington, Carolinas) will employ record-linkage methods that make use of partial identifying information. Duplicate cases will be removed when identified. The specific approaches to duplicate removal are described in the center-specific protocols.

20.3. CASE VALIDATION

It is important that all registered cases have a valid diagnosis of diabetes made by a health care provider. Prevalent 2017 cases of diabetes are validated if they have:

a. been previously validated by SEARCH as prevalent and/or incident diabetes case; or

b. a physician diagnosis of diabetes; or

c. the parent or youth self-reports a physician diagnosis of diabetes on the IPS.

A "physician diagnosis" of diabetes is made if any of the following conditions are met:

    a. a physician diagnosis of diabetes is documented in the person's medical record;

    b. the diagnosis of diabetes is directly verified by a clinician or the person with diabetes is directly "referred" to the study by a clinician;

    c. the person is included in a clinical database that has a requirement for validation of a diagnosis of diabetes by a clinician.

20.4. CASE REGISTRATION

Case Registration closes the gap between case ascertainment, which is a local effort, and data collection, which is centralized at the Coordinating Center. Case Registration should occur after a unique (unduplicated) case is validated and eligibility confirmed based on age, residence/membership, military status, and institutional effort. Information gathered at the time of case registration is minimal. Depending on the local case ascertainment approach, some data elements requested for case registration may be missing at the time of registration and will be completed or verified later through the IPS or core data collection form. For all centers, identifying information such as name, address, phone number, medical record number, etc. will remain at the local center and will not be forwarded/uploaded to the Coordinating Center.

Some of the prevalent 2017 cases, including those who were registered as incident or prevalent cases in earlier years of the study, will already be registered. Sites should ensure that these cases are not duplicative of previously registered cases. Each individual should only have one participant ID number.

Cases may be ascertained and registered up to 22 months after 12/31/2017. Thus, prevalence 2017 will close on 10/31/2019.

20.5. INITIAL PARTICIPANT SURVEY

For 2017 prevalent cases, data collected by the IPS serve two major purposes: a) confirmation of eligibility (especially 2017 residence/membership and date of birth, if data are not available from clinical/administrative records); b) obtain self-reported race/ethnicity (if not already collected by SEARCH).

In addition, data collected by the IPS are also useful for: c) case validation (if not already confirmed); d) determine date of diagnosis (if not already determined); and e) obtain updated contact information (for local use only).

For 2017 prevalent cases, the IPS will be administered to cases that meet the following criteria:

a. all prevalent 2017 cases that are <u>new to SEARCH</u> (which will include all incident 2017 cases that reside in the prevalence area*);

b. all prevalent 2017 cases who were previously registered but have not completed the IPS if they have not previously explicitly refused to complete an IPS (each site is responsible for ensuring compliance with local IRB requirements).

   * The incidence and prevalence areas are the same for California, Washington and Ohio and different for Carolinas and Colorado.

20.5.1. *Introductory Letter*

The Introductory Letter describes the purpose of the study to potential participants ≥ 18 years of age and to the parent/guardians of youth with diabetes < 18 years of age at the time of contact. Each center will customize their letter according to their local operations and IRB requirements. Letters to participants ≥ 18 years of age will be addressed to the Participant. Letters to participants < 18 years of age will be addressed to their parent or guardian.

20.5.2. *Initial Participant Survey*

The IPS may be completed either as a self-administered form, a telephone interview by study staff or a survey research company, online, or during an in-person interview. This flexibility maximizes the likelihood of completing the IPS and obtaining a minimum amount of data for each study participant. The content of the IPS is identical regardless of the mode of administration. Mode of administration will be recorded.

20.6. EXTENDED CORE FORM

In order to address the primary aim of this study, i.e., estimation of prevalence of diabetes by diabetes type, age, sex, and race/ethnicity, it is critical to have core information related to participants' diabetes type as well as their demographic information. These data are obtained using the Extended Core Form (ECF) as well as other sources. The ECF is to be completed for all Prevalent 2017 Cases for whom an ECF has not previously been completed.

The main purpose of the Extended Core Form (ECF) is to collect updated information on most recent assignment of diabetes type (by the health care provider). In addition, the ECF can be used to collect any other missing "core" data, which were not available from clinical/administrative searches or the IPS (date of birth, sex, date of diagnosis, race/ethnicity). If the ECF has been completed previously (for 2017 prevalent cases registered in early years of the study), it does not need to be completed again. However, the diabetes type for 2017 will need to be identified and recorded.

Additionally, for all 2017 prevalent cases, the SEARCH study will attempt to obtain information about type of diabetes based on the physician's most recent report, preferably from 2017. No attempt will be made to obtain other information from the medical record. If multiple assignments of diabetes type are available throughout the medical record, the one that was recorded most recently will be collected (so, if a SEARCH case had a diagnosis type recorded in 2010 and a different one in 2016, the 2016 diagnosis type will be recorded). We are interested in the most recent assignment of diabetes type for several reasons: a) to replicate the 2001 and 2009 study; and, b) to capture possible temporal changes in assignment of diabetes type by provider, as a result of increased availability of relevant information, or changes in diagnostic norms or criteria.

Diabetes type will be sought on all 2017 prevalent cases, including those registered in previous years of the study. Depending on center-specific sources of case ascertainment, most recent assignment of diabetes type can be obtained from the following sources:

a. provider's case reports made during the process of case finding or case validation;

b. clinical database searches, during the process of case finding or case validation (if type in the database is updated on a regular basis);

c. medical record review, during the process of case validation, for cases new to SEARCH and for all other prevalent 2017 cases that are not also incident in 2017.

20.7. ALL PREVALENT 2017 IDENTIFICATION FORM

The All Prevalent 2017 Identification Form will be completed and data entered for all eligible 2017 cases. The purpose of this form is to record relevant case information in the central database for all prevalent 2017 cases. This information includes:

- confirmation of prevalent case

- date identified as a prevalent case

- diabetes type in 2017

- date of diabetes type

- eligible county (geographic sites) - will not be shared with CoC (optional)

- eligible zip code (geographic sites) - will not be shared with CoC (optional).

Note that the Prevalent 2017 Identification Form, a flow chart indicating required data collection for Prevalent 2017 cases and a bulleted checklist are included as Appendix 2.

20.8. UPDATING CONTACT INFORMATION

Attempts will be made to contact the participants to update their contact information (address, phone number, email address, etc.). Sites will use their usual strategies to obtain

this information. This information will be entered in the Tracking Database and maintained locally.

20.9. ASSESSMENT OF THE COMPLETENESS OF CASE ASCERTAINMENT

The validity of prevalence estimates from SEARCH is critically dependent on complete ascertainment of cases through the case-finding approaches described above. An attempt to assess the completeness of case-ascertainment is thus crucial to the SEARCH objectives. The ideal way to determine the completeness of case ascertainment would be 100% review of every medical record in a geographic area or in a health plan to determine if a valid case exists. This requires resources beyond those available for the SEARCH study.

20.9.1. *Geographic Centers - Capture-Recapture*

Capture-recapture (C-RC) is a statistical approach that attempts to estimate the completeness from incomplete samples. This method requires a minimum of two data sources in which a case can potentially be identified. The data elements that are required to derive estimates of the completeness of case ascertainment using capture-recapture methods are 1) the source(s) of the case record for each unique case identified and 2) the date of inclusion in the data source. C-RC analysis is based on cases after deduplication. The best statistical methods will be used, incorporating multiple ascertainment sources, with adjustment for non-independence of data sources. C-RC methods will be used in the geographically-based SEARCH centers with multiple sources of cases (Ohio, Colorado, Washington, Carolinas). Please see the C-RC technical report for additional information about this method.

20.9.2. *Membership Based Center*

Completeness of case ascertainment is also crucial in the membership-based center (California). Since the clinical and administrative data used to ascertain cases are not independent, the capture-recapture approach cannot be applied to these centers. However, every effort should be made to identify all potential cases and then to validate these cases against the SEARCH case definition.

20.10. DENOMINATOR ESTIMATION

Overview

To estimate prevalence accurately, it is necessary to align the numerator and the denominator. That is, the cases of diabetes that are counted in the numerator for prevalence must derive from the same population that comprises the denominator. To estimate prevalence accurately, it is also necessary to have accurate information on the population of children 0-19 years of age from which the diabetes cases are ascertained. The numerator and denominator will be aligned by applying the same criteria for inclusion in the numerator (cases) and the denominator (underlying population) for the study.

The four geographically-based centers (Ohio, Colorado, Washington, Carolinas) will use the US Census non-institutionalized non-military resident population in the area from which cases are drawn as the total denominator for estimation of prevalence. For the 2017 prevalence component of the study, the most recent intercensile US Census data (most likely 2015) will be used. A resident is defined by the Census as a person with a permanent address within the defined geographic area at any time in the index year who is not known to be living elsewhere or only temporarily residing at the eligible address.

The California site will use administrative data to identify members' ages 0-19 years in the participating health plans on 12/31/2017 as the total denominator for estimation of prevalence. The decision not to use membership on any day, which includes those with only one day, was made after several discussions when the 2001 analysis was done.

Special populations

College students are counted in the Census in their residence location as of April 1. This will usually be the college/town where they attend school. An attempt will be made to identify diabetes cases in age-eligible college students resident in geographic areas of SEARCH case ascertainment so that numerator and denominator will be aligned. Youth who are attending college and are members of the participating health plan cannot be identified as attending college using administrative health plan data. Thus, they will be included in the denominator for the member-based centers. Thus, numerator and denominator remain aligned. The handling of college students is also consistent between the geographically based and membership-based center. Thus, numerator and denominator remain aligned. The handling of college students is also consistent between the geographically-based and membership-based center by including these youth in both the numerator and denominator.

Military personnel are counted in the Census at the base/community where they are assigned. Initial total population estimates include these persons. However, as the Census results are further refined, the military personnel are identified separately. No attempt will be made to identify diabetes cases in active-duty military personnel. Thus, final numerator and denominator estimates will exclude active-duty military service members.

Military dependents are counted in the Census at their usual residence, whether on or off base. Thus, they will be counted in the non-military denominator. Medical care for dependents will differ between base locations and access to care systems (military or civilian) will determine the ability to identify cases. Every attempt will be made to identify such cases in a consistent way across centers to align numerator and denominator similarly across centers. Center specific documents define how each center will identify cases in the military dependent population.

Institutionalized persons living in prisons, chronic care hospitals, and other institutions are removed from the counts of the civilian, non-institutionalized denominator, and will not be eligible as cases to align numerator and denominator.

Age, Gender, and Race/Ethnicity: Prevalence

For the geographically based centers (Ohio, Colorado, Washington, Carolinas), census data by age, gender and race/ethnicity will be used to estimate the number of persons in the denominator by age, gender and race/ethnicity. The most recent vintage file containing the intercensile population estimates (most likely 2015) will be used to calculate age, gender, and race/ethnicity specific denominators.

For the California center, direct counts of members 0-19 years by one-year age increments and by gender will be obtained from membership records. Then, the addresses for these members in 2017 will be geocoded to the census block level. Race/ethnicity by one-year age increment and gender will be estimated based on the racial/ethnic composition of these census blocks that are derived from the most recent data available (most likely 2015). This is consistent with methods in earlier years of the study.

For all centers, race/ethnicity data will be collapsed into the following groups (Non-Hispanic White, Hispanic, African American, Asian, Pacific Islander, American Indian/Native American, Other, Multiple, and Unknown) using rules and conventions developed by the Census.

Appendix 1: Site-Specific Case Ascertainment Methods

Methods: Case Ascertainment - Kaiser Permanente Southern California

Kaiser Permanente Southern California (KPSC) is a managed health care organization that provides health care to over 4 million members in southern California. Members with diabetes from seven of the eight counties in the KPSC region have been included in SEARCH since the beginning of the study. The eighth county, San Diego, was excluded due to the geographic distance from the main research office.

GOAL

The goal in the California center is to ascertain and validate all unique (non-duplicated) cases of diabetes in youth aged 0-19 years that are members of KPSC in the seven counties comprising the surveillance area for the SEARCH study to allow for the estimation of prevalence by age, gender, diabetes type and race/ethnicity.

1. CASE ASCERTAINMENT

   1.1. Site-specific case finding approaches: **The California center will use a multi-stepped approach to identify prevalent cases in 2017.**

      1) The first step will be to query the multiple electronic data systems used for cases ascertainment (inpatient and outpatient diagnosis files, including those from outside claims; pharmacy files; and laboratory files) to identify individuals who may have diabetes. This will be done in the same manner used to identify incident diabetes cases. The list of MRNs generated from this linkage will be checked against two lists - those who have previously been registered for the SEARCH study (see step 2) and those who have previously been identified as having diabetes but who did not meet the eligibility criteria to be an incident case (see step 3). Individuals included in step 2 (previously registered cases) and step 3 (individuals known to have diabetes but not registered) will be put in a file and dealt with separately. Staff will then move on to the following steps:
         a) This list will include the individuals MRN and the indicator(s) that identified them, as possibly having diabetes, the medical records for these individuals will be reviewed for a physician's diagnosis of diabetes (validation) as well as eligibility per standard SEARCH protocol. This step will be facilitated by an application similar to NLP, which will pull key information out of the EHR for easy review. Staff members continue to have access to the EHR as needed.
         b) These cases will include those who are incident in 2017 and those who are prevalent in 2017 but were not previously registered or identified as having diabetes.
         c) The majority of the "prevalent only" cases will have joined the health plan after their diabetes diagnoses. If we identify cases that were diagnosed when they were a KPSC member in earlier years but were missed in previous cycles of cases ascertainment, they will be registered as prevalent 2017 cases as well as an incident case for the appropriate incident year (if permitted by the study rules).
         d) This process will be repeated at least three times per year until the registration period for prevalent 2017 cases is closed.
         e) Diabetes type in 2017 will be recorded for all of these cases.
         f) We will attempt to contact all of these cases and invite them to complete the IPS.

2) As noted in step 1, the second step will be to link the list of MRNs generated in Step 1 to the list of all SEARCH California cases registered in calendar years 2002-2016, which included 2001 and 2009 prevalent cases and incident cases in 2002-2016 using their KPSC unique MRN. This will allow for quick identification and deduplication of previously registered cases without requiring a complete chart review for validation. Eligibility for the prevalent cohort will then be assessed.
   a) Those who are not members anytime during calendar year 2017 (1/1/17-12/31/17) will be excluded (will not be identified as prevalent 2017 cases).
   b) Those who remain members will have their birthdate queried to determine if they will be 20 years of age before 12/31/2017. Those who will turn 20 before the end of 2017 will be excluded.
   c) Those who will not be 20 by the end of 2017 will have their addressed in the membership files queried to determine if they live in the surveillance area. Those whose address is in San Diego County or in any other county outside of our surveillance area will be excluded.
   d) All of the remaining cases will be identified as being prevalent 2017 cases by selecting that option in the tracking database and uploading them to the Coordinating Center.
   e) Medical records for 2017 will be reviewed to determine diabetes type in 2017. The diabetes type in 2017 will be entered in the TDB and uploaded to the Coordinating Center.
   f) Individuals that did not complete an IPS in earlier years of the study because they could not be contacted will be sent a letter requesting that they complete an IPS. Letters will be sent to parents/guardians of cases < 18 years of age. Those who refused previously or reached the maximum number of contact attempts will not be re-contacted.

3) The third step will be to link the list of MRNs generated in Step 1 to the list of individuals known to have diabetes (valid cases) who were identified in previous (non-prevalent) years who did not meet the case definition for an incident cases. This information has been retained for the last 3 years.
   a) Those who are not members anytime during calendar year 2017 (1/1/17-12/31/17) will be excluded.
   b) Those who remain members will have their birthdate queried to determine if they will be 20 years of age before 12/31/2017. Those who will turn 20 before the end of 2017 will be excluded.
   c) Those who will not be 20 by the end of 2017 will have their addressed in the membership files queried to determine if they live in the surveillance area. Those whose address is in San Diego County or in any other county outside of our surveillance area will be excluded.
   d) A core from will be completed which will include diabetes type in 2017 as well as other core variables.
   e) These cases will be registered as prevalent 2017 cases by selecting that option in the tracking database and uploading them to the Coordinating Center.
   f) We will attempt to contact all of these cases and invite them to complete the IPS.

1.2. Identification of duplicate cases

The SEARCH tracking database (TDB) is designed to prevent the entry of more than one record for any individual due to the uniqueness of their MRN. On rare occasions, a health plan member is assigned more than one MRN. To prevent duplicate case entries, all cases are searched and reviewed in the TDB by MRN, first and last name and date of birth (DOB) prior to uploading (registering) new participants. If a case is found with duplicate MRNs, these records are reviewed to confirm that they are duplicates. True duplicate records are not registered, only the original case MRN is retained in the TDB. Health plan administration is notified of any members that were assigned two MRNs to prevent future duplications. If the MRN is changed by health plan administration, then the TDB is updated by SEARCH staff.

1.3. Completeness of case ascertainment

Potentially eligible children and youth are identified through two approaches:

1) Children and youth < 20 years of age at the time of diabetes diagnosis are reported to the SEARCH project manager monthly by the Kaiser Permanente Pediatric Endocrinologists, and
2) Queries of KPSC electronic data systems including the electronic health record are queried 3-4 times a year to identity children and youth <20 years of age with ICD-9-CM or ICD-10-CM codes for diabetes and associated conditions (inpatient and outpatient), pharmacy dispenses of drugs to treat diabetes, and laboratory test results indicative of diabetes.

By obtaining case reports from Pediatric Endocrinologists or their nurses monthly (with reminders from the Project manager if no cases are reported), updating our list of Pediatric Endocrinologists whenever new physicians are added to the group or leave and by querying the electronic data systems 3-4 times a month for individuals with any indication of diabetes in the past 12 months, we believe that the completeness of case ascertainment is high. Since KPSC is an integrated health care system, we do not have independent sources of cases that can be used for traditional capture-recapture analysis.

Methods: Case Ascertainment: Colorado - SEARCH for Diabetes in Youth

GOAL

The goal is to ascertain and validate all unique (non-duplicated) cases of diabetes in youth aged 0-19 years who reside in the Colorado Center prevalence areas in 2017. This will allow estimation of prevalence rates by age, gender, and ethnicity.   The "Colorado Center" prevalence areas include: a) specific counties in Colorado; and b) the Navajo Nation reservation in Arizona and New Mexico.

Site specific methods for case ascertainment, validation and confirmation of eligibility will mirror those employed in the 2001 and 2009 prevalence studies.

1. DENOMINATOR ESTIMATION

    1.1. Site - specific approaches

        a) Colorado Locations

        The 2010 US Census non-military resident population from which cases are present in the index year will be used as denominator. A "resident" is defined as a person with a permanent address within the defined geographic area at any time in the index year, who is not noted to be living elsewhere and only temporarily residing at the eligible address.

        Active duty military individuals are not eligible to be counted in the numerator. However, military dependents that have access to civilian medical facilities will be captured in the numerator. Those that use military facilities will be excluded. Military personnel and dependents are counted by the US Census in the denominator for the county in which they currently reside. To align the numerator with the denominator, active duty military individuals need to be subtracted from the denominator. The number of Active Duty Military (ADM) that needs to be subtracted from each denominator cell (age, race, gender) will be estimated based on counts of ADM by age provided by the Military Family Resource Center (MFRC), and based on the racial and gender distributions of military personnel for our desired geographic area provided by the Census Bureau. The racial distribution will be calculated from Census Bureau totals of a cohort of ADM between 16-18 years old (thus excluding 19 year olds that SEARCH will collect). The racial distribution of this population should not vary much from the racial distribution of the 17 to 19 year olds. The gender distribution will be derived from a much larger age range, 18 to 64 year olds. In all of ADM nationwide, females account for 15% of that population. This is consistent with the proportions that we have calculated for our incident and prevalent areas. Once all of these proportions (by age, race and gender) are applied to the counts from MFRC, estimates of ADM will be subtracted from the Colorado denominator. There is one military base in Denver: Buckley Air Force Base.

        b) Navajo Nation Reservation

        We will use the Indian Health Service (IHS) user population for eligible service units on the Navajo Nation, defined as persons with one or more visits in the past 3 years (including the index year). For example, the 2016 prevalence year includes 2014,

2015, 2016 users of IHS facilities. This is consistent with methods used by the IHS for other research conducted at their facilities. There are small differences between census and IHS denominators, which will be the source of numerator cases. Census denominators often include persons not otherwise eligible for IHS care.

1.2. Denominators for prevalence

Prevalence of diabetes in youth aged 0-19 will be estimated for the year 2017 in the following geographic areas: Colorado Urban-suburban counties, including the Denver-Boulder metropolitan statistical area, (Denver, Adams, Arapahoe, Broomfield, Douglas, Jefferson, and Boulder).

Rural Colorado counties, including the San Luis Valley (Conejos, Costilla, Alamosa, Saguache, Mineral, Rio Grande) in south-central Colorado, and Mesa county in western Colorado. Selection of these counties was based on several reasons: a) they participated in 2001 and 2009 prevalence studies; b) a network of collaborators, health care providers, hospitals, etc., has already been developed through several previous studies in the San Luis Valley region; and c) their participation will help investigators learn about the ascertainment process in these areas, patterns of access to different sources of cases, patient referral patterns, and issues of confidentiality in different populations and provider groups.

Navajo Nation in Arizona and New Mexico-8 Service IHS Units including Shiprock, Kayenta, Chinle, Winslow, Tuba City, Fort Defiance, Crownpoint, and Gallup.

2. PREVALENT CASE ASCERTAINMENT

2.1. Case finding site - specific approaches

Data sources

Cases will be identified through multiple approaches, which are site and area dependent. These are tabled in **Appendix 1** for each of the areas included. In each area, multiple sources will be used to ensure that as few cases are missed as possible.

- Colorado Locations

  The types of data sources include: pediatric endocrine clinical computerized databases, HMO computerized diabetes registries, primary care practices charts, private practices charts, computerized and non-computerized hospital records, diabetes educators case records. A core network of providers identified through response to a statewide survey has been developed. In most situations, possible cases will be identified through database searches (see description in Appendix 1). Chart reviews, hospital record reviews from selected hospitals, letters and telephone surveys to primary care practices, and fax update reminders are other additional methods that will be used to ensure as a complete ascertainment of cases as possible.

- Navajo Nation

  Eligible cases will be identified essentially through on source: the Indian Health Service computerized hospital and ambulatory database (RPMS).

IRB considerations

Based on our prior experience with the 2001 and 2009 prevalence studies, in most instances, we will be able to identify potential cases without prior consent from the patient through an IRB approved HIPAA waiver. IRB review and approval of procedures for the 2017 prevalence study will occur before any case ascertainment begins. IRB approval will be requested from each institution involved in the study, and from the Navajo Nation.

Identification of duplicate cases

a) Colorado Locations:
   Management of duplicates will occur electronically as well as manually. Once personal identifiers (name, gender, date of birth, ethnicity, zip code, etc.) are obtained, possible duplicates will be checked electronically/manually and all duplicate cases will remain in the SEARCH Tracking Database as a means of assessing ascertainment completeness-but they will not be registered or counted. Each unique case will be registered as identified from the first source reporting the case. Cases that cannot be determined to be unduplicated will be marked for further data collection if possible, by contact with parent or case.

b) Navajo Nation:
   Since case ascertainment involves primarily one source (the IHS), there is no need for matching for duplicates across sources. Matching for duplicates within the main IHS sources will occur internally (i.e., conducted by the SEARCH - Navajo coordinator).

3. CASE VALIDATION AND CONFIRMATION OF ELIGIBILITY

   3.1. Site-specific methods

The methods to be used by the Colorado site are outlined in Appendices 1 and 2 for each source of cases. For cases new to SEARCH and identified through administrative database searches, validation will occur through medical record review for a physician diagnosis of diabetes. If the prevalent case has been previously identified by SEARCH, or is referred to the study by a clinician, there is no need for further validation.

a) Colorado Locations
   Specific information about the geographic area of residence in 2017 such as street address and zip code is needed to determine eligibility and residence during the prevalence year. As permitted by various IRB's, cold calling of potentially eligible individuals will be performed in order to confirm eligibility through completion of the Initial Patient Survey Part 1 over the phone.

b) Navajo Nation
   Information about tribal affiliation, Native American ethnicity, and use of IHS facilities in the past 3 years in needed to determine eligibility. Such information is typically

obtained through medical record abstraction during the process of validation, and therefore will not involve any contact with the prevalent case.

4. COMPLETENESS OF CASE ASCERTAINMENT

Capture-recapture methods will be used to calculate the completeness of case ascertainment. The best statistical methods will be used, incorporating multiple ascertainment sources, with adjustment for non-independence of data sources.

Capture-recapture can only be used when multiple sources of cases exist. Therefore, we will be able to assess completeness of ascertainment for the Denver-Boulder Colorado locations, but not for the Mesa County location, for the San Luis Valley locations, or for the Navajo Nation.

Data elements required for calculation of capture-recapture estimates are:

- source of case record
- date of inclusion on data source
- record numbers to remove duplicates from same data source
- PHI's to identify duplicates between different data sources.

APPENDIX 1 - Sources of Prevalent Cases SEARCH

Colorado counties: Denver, Adams, Arapahoe, Douglas, Jefferson, Broomfield, Boulder. For capture-recapture estimates, each unduplicated case will record all the sources in the seven counties where it was identified.

Source	Type	Record system	Case finding	Validation
Barbara Davis Center (BDC) for Childhood Diabetes	Pediatric Endocrinology	Computerized clinical database	Search of clinical computerized database (excluding codes for "rule out" diabetes; sibling); ICD codes not used in database	Not needed
Denver Health	HMO/ Community Health Center	Computerized diabetes registry	Diabetes registry using computer algorithm: ICD codes 250.XX, include neonatal diabetes (775.1), and 362.0X (diabetes retinopathy). Exclude: gestational diabetes (648.8); hyperglycemia NOS (790.6), nonclinical diabetes (790.2).plus eligibility criteria	Denver Health chart review
Pediatric endocrinology practices	Private practice	Non-computerized and computerized medical records	Chart review	Not needed
Primary care practices from Colorado provider list by type of practice and geography	Primary care	Mixed manual and computerized	Case report form completed by provider and given to SEARCH clinic	Not needed

Source	Type	Record system	Case finding	Validation
Hospitals-other than Community Hospitals -Exempla -Centura -Children's	Hospital	Computerized and non-computerized administrative records	Query run on database using computer algorithm: ICD codes 250.XX, 362.0X (diabetes retinopathy), and 775.1 (neonatal diabetes). Exclude: gestational diabetes (648.8); hyperglycemia NOS (790.6), nonclinical diabetes (790.2) plus eligibility criteria.	Chart review

San Luis Valley area counties: Conejos, Costilla, Alamosa, Saguache, Mineral, Rio Grande. No capture-recapture estimates can be done in this area, due to insufficient numbers of sources.

Source	Type	Record system	Case finding	Validation
Community Health Centers:(Valley Wide Health Systems)	Federally funded community health center for primary care	Computerized billing database	Diabetes registry with chart augmentation Diabetes registry using computer algorithm	Chart review
Diabetes educators	Clinician referral	Case records	Certified Diabetes Educators complete and return Case Report Form to SEARCH clinic	Not needed

Mesa County (Grand Junction area): No capture-recapture estimates can be done in this area, due to insufficient numbers of sources

Source	Type	Record system	Case finding	Validation
Community Hospitals: St Mary's; Community Hospital	Hospital	Computerized and non-computerized records	Query run on database using computer algorithm: ICD codes 250.XX; 362.0X (diabetes retinopathy), 775.1 (neonatal diabetes). Exclude: gestational diabetes (648.8); hyperglycemia NOS (790.6), nonclinical diabetes (790.2) plus eligibility criteria.	Chart review
Diabetes educators	Clinician referral	Case records	Certified Diabetes Educators complete and return Case Report Form to SEARCH clinic	Not needed

APPENDIX 2 - Native American Sites

NAVAJO NATION

Source	Type	Record system	Case finding	Validation
Indian Health Service data base	Ambulatory care and hospitals	Computerized hospital and ambulatory record system	Data base-RPMS-logic Developed ICD codes 250.xx and 775.1 (neonatal diabetes)	Chart review

Methods: Case Ascertainment - Children's Hospital Medical Center, Cincinnati, Ohio

Cincinnati Children's Hospital Medical Center (CCHMC) is the only pediatric healthcare facility serving southwest Ohio, northern Kentucky, and southwest Indiana. As a result, children and adolescents with complex medical problems are referred almost exclusively to CCHMC. The majority of patients served by CCHMC are residents of one of eight counties surrounding the hospital. These eight counties make up the primary service area for the hospital.

GOAL

The goal in Cincinnati is to ascertain and validate all unique (non-duplicated) cases of diabetes in youth aged 0-19 years in the 8-county primary service area for CCHMC. This will allow estimation of prevalence by age, gender, and ethnicity.

1. CASE ASCERTAINMENT

   1.1. Case finding site - specific approaches

   Primary Source for Case Identification

   The primary source for identification of prevalent cases will be a list of patients who attend Diabetes Clinic at CCHMC. This list will be generated via EPIC.

   Expanded Sources for Case Identification

   Expanded case identification will need to be done to identify subjects who meet the eligibility criteria, yet who have never been seen at CCHMC. We have established a network to identify these cases:

   a) Hospitals (total of 11)
   b) Bureau of Children with Medical Handicaps (BCMH)

   1.2. Identification of duplicate cases

   The SEARCH Tracking Database is designed to prevent the entry of more than one record for any patient due to the uniqueness of the medical record number. Occasionally, however, a patient is mistakenly assigned more than one medical record number, thereby permitting a duplicate entry in the SEARCH Tracking Database. Each month a query is run to identify patients in the Database who have matching entries for both date of birth and last name or date of birth and first name. This method allows two opportunities to identify duplicate records. When duplicate records are identified, these records are reviewed to confirm that they are duplicates. True duplicate records are marked as duplicate records.

   For expanded sources of case identification, we will identify duplicates by comparing names and dates of birth within our SEARCH Tracking Database to the lists provided by our network partners.

   1.3. Completeness of case ascertainment

   All patients who present to CCHMC for diabetes care are identified for inclusion in the SEARCH Tracking Database using two independent methods.

a) All follow-up visits to the Diabetes Clinic are entered into the SEARCH Tracking Database on a monthly basis.

b) Twice each year a list is generated of all inpatient, ER, and outpatient visits who were assigned any of the following ICD-10 codes: E08.xx, E09.xx, E10.xx, E11.xx, E13.xx, P70.2, and O24.1. This list is compared to the SEARCH Tracking Database, using a unique medical record number. All cases that appear on this monthly list, but are not included in the database, are marked for review to determine SEARCH eligibility.

Sensitivity for case finding within CCHMC is very high due to the independence of the two methods described above.

We will use our network partners (hospitals and BCMH) to maximize the sensitivity of our expanded case-finding outside CCHMC. The source(s) of each validated case, as well as the type of visit (inpatient, ER, or outpatient) will be recorded in the SEARCH Tracking Database. This information will then be used to determine the capture-recapture estimates for completeness of ascertainment.

Methods: Case Ascertainment - Washington

GOAL

To ascertain and validate all unique (non-duplicated) cases of diabetes in youth aged 0-19 years in the Puget Sound region of Washington. This region includes 5 counties: King, Kitsap, Pierce, Snohomish, and Thurston. This work will allow estimation of prevalence 2017 rates by age, gender, and race/ethnicity.

1. DENOMINATOR ESTIMATION

   1.1. Site specific approach

   The 2010 US Census non-institutionalized, non-military resident population of the 5-county Puget Sound Region will be used. Projections beyond 2010 will be made by the WA State Office of Financial Management (OFM) that provides the executive branch, the legislature, and the public with estimates, forecasts, and reports on the state's population. Various public and private organizations rely on data developed and maintained by OFM for planning and assessment purposes. The agency also serves as a liaison with the federal Census Bureau.

   1.2. Special populations

   College students: An effort will be made to contact medical facilities affiliated with larger institutions in this region, e.g., University of Washington, to find cases since college students will be included in the denominator.
   Military personnel: The pediatric endocrinology group affiliated with Madigan Army Medical Center will continue to refer cases (non-military personnel, i.e., children of military personnel) to SEARCH.

   Native American residents: There are several Native American reservations in the Puget Sound region (including Swinomish, Tulalip, S'Klallam Port Gamble, Port Madison, Puyallup, Muckleshoot, Nisqually Indian Reservations) as well as many urban, non- reservation based American Indians. Given the large percentage of urban Indians in our area and the likely referral of American Indian youth with diabetes to non-reservation based clinical sites, local case ascertainment will not focus on reservation-based youth.

   1.3. Denominators for prevalence

   For the prevalence estimate, projections based on the 2010 US Census made by the WA State OFM will be used. The estimated populations residing in the Puget Sound Region is summarized in Table 1.

Table 1. Estimated prevalence denominators by age group and ethnicity*

Age Group	White	Black	AIAN	Asian	NHOPI	Two or More Races	Total
0 - 4	171,138	17,085	4,150	26,112	3,207	31,855	253,547
5 - 9	168,992	15,898	4,019	26,174	2,925	26,698	244,706
10 - 14	175,116	16,067	4,115	24,436	2,955	24,189	246,878
15 - 19	184,022	16,708	4,572	28,520	3,222	20,795	257,839

*Based on 2010 census estimates [WA State Office of Financial Management]

2. CASE ASCERTAINMENT

2.1. Data sources

A combination of clinical referrals and non-clinical or administrative data sources will be used to identify prevalent cases.

a) Pediatric endocrinologists

There are 4 pediatric endocrinology groups in the Puget Sound region, all of which have been referring cases to SEARCH to date. They include:

- Seattle Children's
- Mary Bridge Children's Health Center
- Pediatrics Northwest, Tacoma
- Madigan Army Medical Center Diabetes Care Center

b) Adult endocrinologists

Adult endocrinology practices that have been involved in local SEARCH case ascertainment to date include:

- The University of Washington Medicine Diabetes Care Center
- Swedish Diabetes Education Center

c) Hospitals

Two major pediatric hospitals serve the 5-county Puget Sound area:

- Seattle Children's
- Mary Bridge Children's

These two pediatric hospitals have been referring cases to SEARCH. Other area hospitals with a history of providing care to youth with diabetes will also be included:

- Harborview Medical Center
- Kaiser Permanente
- Providence St. Peter Hospital

- Virginia Mason Medical Center
- Swedish Medical Center

d) Primary care clinics and other sources

Cases have also been referred by smaller clinics and by direct contact with participants, including:

- SeaMar Community Health Centers
- University of Washington Physicians
- Seattle Children's Odessa Brown Children's Clinic
- Community Health Centers of King County
- Neighborcare Health)
- Diabetes-related events such as camps, walks, or expos
- Public website, social media, or advertisements

2.2. Identification of prevalent cases: determining eligibility and deduplication

Provider referrals and medical records will be used to identify potentially eligible youth.

SEARCH staff will use available data to compare and deduplicate cases prior to registration. As needed, SEARCH staff will ask participants to complete surveys online, via phone, or by mail.

The following items will be considered as possible matching variables, depending on the information available:

- Name
- Gender
- Date of birth
- Race/ethnicity
- Geographical information – address and/or ZIP code
- Social security number
- Medical record numbers
- Telephone number(s)
- Email addresses
- Admission date of hospitalization(s)
- Parent/guardian names
- Partial name matches

All sources used to identify cases (e.g. clinical records, hospital records, registries, etc.) will be recorded. We anticipate that most cases will be identified by at least 2 distinct sources (e.g. clinic and hospital record). We will use a combination of manual, deterministic, and probabilistic record linkage methods to remove duplicates.

2.3 Case definition and eligibility

a) Prevalence: The case must have diabetes during 2017.

b) <u>Date of diagnosis</u>: Is the date of first clinical diagnosis of diabetes in a non-pregnant state, as validated via provider referral, self-referral, or medical record review.

c) <u>Date of birth</u>: Age less than 20 years on December 31, 2017; this corresponds to a birth year: 01/01/98 – 12/31/2017 for the prevalence year of 2017. Cases that reach age 20 years during the prevalence year are not eligible.

d) <u>Age range</u>: 0 – 19.999 in the index year.

e) <u>Geographic area</u>: 5-county Puget Sound Region, Washington (King, Kitsap, Pierce, Snohomish and Thurston Counties).

f) <u>Resident of population</u>: Non-institutionalized, non-military resident or member of population from which cases are present in the index year.

g) <u>Military personnel</u>: Active duty military personnel will be excluded to align with the denominator. Military dependents that have access to civilian medical facilities will be identified as cases in the numerator and will be included in the non-military denominator.

h) <u>College students</u>: College students are eligible, as they will be included in the non-institutionalized, non-military denominator.

3. CASE VALIDATION

3.1. Site-specific methods

All cases will be confirmed to have provider-diagnosed diabetes prior to case registration. The majority of prevalent cases will be validated by provider verification since cases will primarily be identified and approached through the hospital or endocrinology practices.

Validation sources may include:

- Medical record review for provider diagnosis of diabetes
- Direct verification of case status by health care provider
- Location in clinically verified database (where case has been verified by a clinician)
- Interview of parent(s) or self-report of physician diagnosis of diabetes via IPS

4. COMPLETENESS OF CASE ASCERTAINENT

4.1. Capture-recapture

Capture-recapture methods will be used to calculate the completeness of case ascertainment. The best statistical methods will be used, incorporating multiple primary ascertainment sources (e.g. outpatient, hospital, etc.), with adjustment for non-independence of data sources.

Appendix 2: Prevalent 2017 Forms / Flow Chart / Checklist

Patient ID Number

Site	Sub-site	Sequential ID

All Prevalent 2017 Identification Form

1. Prevalent 2017 Confirmed: ☐

2. Date Captured: ☐☐ ☐☐ ☐☐☐☐
   Month   Day   Year

3. Provider Type (closest to 2017):

| 1☐ Type 1 (IDDM) |
| 2☐ Type 1A |
| 3☐ Type 1B |
| 4☐ Type 2 (NIDDM) |
| 7☐ Other (specify): |

Code: ☐☐☐☐

4. Date of Provider Type: ☐☐ ☐☐ ☐☐☐☐
   Month   Day   Year

OPTIONAL – NOT FOR DATA ENTRY

5. Zip Code (2017) ☐☐☐☐☐

6. County and State of Residence (2017)

County [   ]   State [   ]

Prevalence of Diabetes in Youth in 2017

GOALS:

The goal of the "2017 Prevalence Study" is 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 prevalence areas/health plan/IHS in 2017 in order to:

- Estimate the population prevalence of diabetes in youth age < 20 year in 2017 by age, sex, race/ethnicity and diabetes type

- Compare the prevalence of diabetes in youth age < 20 years in 2001, 2009 and 2017, by age, sex, race/ethnicity and diabetes type.

ELIGIBILITY:

- Diabetes (other than gestational diabetes only) diagnosed during or prior to 2017. Medical record(s) will be reviewed to identify diabetes type closest to 2017 (either before or after).

- Age less than 20 years on December 31, 2017; this corresponds to a birth year: 01/01/98 – 12/31/2017 for the prevalence year of 2017. Individuals reaching age 20 years during the prevalence year are not eligible

- Resident of the population defined for prevalent cases at any time during the prevalence year (for geographically-based centers) or member of the participating health plan or IHS region at any time during the prevalence year

  Note: The incidence and prevalence areas are the **same** for California, Washington and Ohio and are **different** for Carolinas and Colorado.

- Not active-duty military

- Not living in an institution (defined as such by the Census).

CASE FINDING:

- Site specific - All of the data sources that were used to identify cases prevalent in 2001 and 2009 and incident in 2002-2017 will be used to identify prevalent 2017 cases if they are still available.

- For geographically based centers, database searches for potential cases will include the three years prior to 1/1/2017, retrospective to January 1, 2014.

- **DO NOT** use the SEARCH database as a source for cases.

- Consult "Site-Specific Case Ascertainment Methods" (Appendix) for your site for further details.

- Cases may be ascertained and registered up to 22 months after 12/31/2017. Prevalence 2017 will close on **10/31/2019**.

ALL PREVALENT 2017 IDENTIFICATION FORM

The All Prevalent 2017 Identification Form will be completed and data entered for all eligible 2017 cases. This form includes:

- Date identified as a prevalent cases
- Diabetes type in 2017
- Date of diabetes type.

PREVALENT IPS FORM

The Prevalent IPS form will be administered to cases that meet the following criteria:

- All prevalent 2017 cases that are new to SEARCH (except Incident 2017 cases).

- All prevalent 2017 cases who were previously registered but did not complete an IPS and did not explicitly refuse to complete an IPS when previously contacted. (Please comply with local IRB requirements.)

EXTENDED CORE FORM

- The Extended Core Form is to be completed for all Prevalent 2017 Cases for whom an extended core form has not previously been completed.