SEARCH FOR DIABETES IN YOUTH STUDY

SEARCH FOR DIABETES IN YOUTH STUDY
OMB No. 0920-0904
Revision

OMB SUPPORTING STATEMENT: PART B

Submitted by:
Sharon Saydah, PhD MHS
Project Officer

Epidemiology and Statistics Branch
Division of Diabetes Translation
Centers for Disease Control and Prevention
4770 Buford Hwy NE
Bldg 107
Atlanta, GA 30341
Telephone: (301) 458-4183
Fax: (301) 458-4025
e-mail: ssaydah@cdc.gov

May 8, 2014
Revised July 31, 2014
Revised November 28, 2016

Table of Contents

B.   STATISTICAL METHODS

B.1   Respondent Universe and Sampling Methods

B.2   Procedures for the Collection of Information

B.3   Methods to Maximize Response Rates and Deal with Nonresponse

B.4   Tests of Procedures or Methods to be Undertaken

B.5   Individuals Consulted on Statistical Aspects and Individuals Collecting and/or

List of Attachments

Attachment 1:   Authorizing Legislation: Public Health Service Act
Attachment 2A: 60 day Federal Register Notice
Attachment 2B: Summary of Public Comments and CDC Response to 60
                    day Federal Register Notice
Attachment 3:   SEARCH Study Sites and Coordinating Center
Attachment 4A1:   Medication Inventory
Attachment 4A2a: Initial Participant Survey, Parent, Incident Case
Attachment 4A2b: Initial Participant Survey, Adult, Incident Case
Attachment 4A3:   Physical Examination Form
Attachment 4A4:   Specimen Collection Form
Attachment 4A5a: Initial Participant Survey, Adult, Prevenlant Case
Attachment 4A5b: Initial Participant Survey, Parent, Prevalent Case
Attachment 4A5-6:   Registry Data Collection, Prevalent Cases
Attachment 5:   Uses of SEARCH data and publications
Attachment 6:   Experts Consulted During SEARCH Development
Attachment 7:   IRB Approvals
Attachment 8:   Certificate of Confidentiality
Attachment 9:   Clinical site consent forms including HIPAA waiver
                    language
Attachment 10:   Data use agreements
Attachment 11:   Published article documenting SEARCH methods

Attachment 12:   SEARCH Technical Report Capture Recapture

B.1. Respondent Universe and Sampling Methods

The respondent universe includes all youth less than 20 years in applicable geographic areas and health plans. SEARCH participants are drawn from four geographically defined populations in Ohio, Washington, South Carolina, and Colorado, from health plan enrollees in California, and from Indian Health Service beneficiaries from American Indian populations in Arizona and New Mexico.

Case Ascertainment Processes utilized by the SEARCH 4 Clinical Sites

Ongoing Case Ascertainment: SEARCH 4 will use the reporting network of clinics and health care providers that was established in SEARCH phases 1, 2 and 3 as **(Attachment 11)** the primary approach to case-finding for incident cases of diabetes for the period 2015-2020. Additionally, the case ascertainment approach involves existing validated pediatric diabetes databases, hospital and health care plan databases, and other health care organizations.

Case Validation: Case identification will remain the same. Cases of diabetes will be validated based on physician reports, medical record reviews, or self-report of a physician diagnosis of (non-gestational) diabetes. A physician-diagnosed case of diabetes is established if any of the following criteria are met: (1) medical record review indicating a physician diagnosis of diabetes, (2) the diagnosis of diabetes is directly verified by a physician, (3) the physician referred a youth with diabetes to the study, or (4) the case was included in a clinical database that had a requirement for verification of diagnosis of diabetes by a physician.

Eligibility Criteria: Eligibility criteria will remain the same. As in SEARCH 1, 2 and 3, the study will be confined to children/youth who, in addition to having an onset of physician-diagnosed of diabetes during the index year, are also are < 20 years of age on December 31 of the index year, and 2) are residents of the population defined for geographically-based centers at any time during the index year, or are members of the participating health plan for membership-based centers at diagnosis, and 3) are not active duty military personnel or institutionalized. Protected Health Information (PHI) will be obtained in order to validate and confirm eligibility and uniqueness of cases in keeping with HIPAA and the procedures and approvals required by the local IRB.

De-duplication: Duplicates will be identified using both electronic files and manually, both within and between case sources, using the name or initials, gender, date of birth, ethnicity, zip code, or other available information, in keeping with HIPAA requirements to use the least amount of PHI in conducting research. The number of duplicates identified will be used to estimate completeness of ascertainment with the capture-recapture method among the geographic centers.

Case Registration: Cases that are valid, eligible and unique will be registered by the center with information being uploaded to the Coordinating Center. Names, addresses, date of diagnosis and date of birth are not provided to the Coordinating Center. In cases where duplicates and cases that are not valid or eligible are identified at a later date, they will be unregistered by both the local center and the Coordinating Center. We estimate that the Registry Study will involve information collection from an average of 1,511 incident cases and 776 prevalent cases

Estimating Trends in Incidence: 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 be estimated using Poisson regression. For example, incidence rates adjusted for race/ethnicity, sex and age can be obtained by fitted the following model:

$$\log\left(N_{ijk}\right) = \log\left(D_{ijk}\right) + \beta_0 + \beta_1 R_i + \beta_2 S_j + \beta_3 A_k \text{(Eq. 1)},$$ where N_{ijk} represent the number of cases observed in the ith race/ethnicity group, j^{th} sex, and k^{th} age group. The race/ethnicity variable used in SEARCH has 5 levels [American Indian (AI), Asian Pacific Islander (API), Hispanic, non-Hispanic black (NHB) and non-Hispanic White (NHW)]. Sex has 2 levels and age is typically grouped into 5-year intervals (0-4, 5-9, 10-14 and 15-19 years). Very few youth under 10 are diagnosed with T2D; therefore, incidence of T2D is reported only for the 2 oldest age groups. The Poisson model is fitted using a logarithm link. Diagnostic tests are performed on the fitted model to ensure that its residual distribution satisfy the underlying assumptions. A Poisson-Gamma or negative-binomial distribution will be used instead of the Poisson distribution if estimated dispersion parameter is significantly greater than 1. The unadjusted trend in incidence can be estimated by fitting a Poisson regression of the form:

$$\log\left(N_t\right) = \log\left(D_t\right) + \gamma_0 + \gamma_1 Y_t \text{ (Eq. 2)},$$ where N_t and D_t represent the number of diagnosed cases and the size of the population under surveillance in year Y_t, respectively. The trend in incidence is measured by the parameter γ_1 in this model. Further adjustments can be achieved by including the appropriate covariates in the model represented in Eq. 2. For example, combining Eq. 1 and Eq. 2 into:

$$\log\left(N_{ijk,t}\right) = log\left(D_{ijk,t}\right) + \beta_0 + \beta_1 R_{i,t} + \beta_2 S_{j,t} + \beta_3 A_{k,t} + \gamma Y_t \text{ (Eq. 3)}$$ will allow us to estimate the trend in incidence adjusting for race/ethnicity, sex and age. The variables $N_{ijk,t}$ and $D_{ijk,t}$ represent the number of cases diagnosed out of the total number under surveillance in each subgroup in a specific year.

Table 1.: Detectable rates of change in annual incidence rate					
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
Non-Hispanic Whites	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
Non-Hispanic Blacks	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
Asian/Pacific Islanders	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
American Indian	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

Power to detect trends in incidence: SEARCH 4 will add five additional years of incidence data, thereby providing improved power to detect minute changes in the incidence rate over the 17-year period of data registry. **Table 1** shows the detectable effect size by race/ethnicity and sex across diabetes type with 90% and 80% power, assuming a 5% significance level. The detectable effect size for each power level was estimated using Nam's approach (Nam, 1987). SEARCH 4 will be well-powered to detect changes in incidence trends in NHW, Hispanic, and NHB youth with Type 1 Diabetes (T1D) and Type 2 Diabetes (T2D). For example, the study will have to 90% power to detect changes in annual incidence rates as small as 1.0% in NHW females with T1D and 2.1% in NHB females with T2D. However, it will have limited power to detect those effects in Asian-Pacific Islander and American Indian youth, with the smallest detectable effect size of ~3.6% with 80% power in T2D youths.

Identifying a change point in the T1D incidence rate in NHW youth: Incidence data collected during the 2002-2009 period suggested a linear trend with a constant rate of increase of 2.7% per year. With the accumulation of five more years of data, SEARCH could detect whether trends in incidence have changed, and estimate retrospectively when the change occurred. Simulation studies were performed to assess the power to correctly identify the year corresponding to the change point. The simulation study started with data already available in SEARCH, which was used to fit a Poisson regression model that predicts the number of incident cases per year as a linear function of the incidence year, using the logarithm of the denominator as an offset (Eq. 2). The number of incident cases was then divided by the observed denominator to provide the fitted incidence rate per year. This model provided fitted data for the first eight years with observed data and predicted the incidence rate for the future incident years until 2018. Projected denominators were obtained at the state and county level for each SEARCH site. **Figure 1**. (panel A) shows the observed (blue squares), fitted (red circles between 2002 and 2008) and projected (red circles from 2009 to 2018) incidence rate from 2002-2018. The model is then perturbed to mimic a change point that could occur in 2012, 2013, 2014, 2015, or 2016, respectively. 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. That is, Eq. 2 was revised as:

$$\log(N_t) = \log(D_t) + \gamma_1 Y_t + \gamma_2 I(Y_t > k)$$ (Eq. 4) where $I()$ is an indicator variable that takes the value of 1 if the incidence year was greater than the year of the change point, and 0 otherwise. Four possible "effect sizes" or percent change (5, 10, 15 and 20%) starting at the change point year were simulated. **Figure 1.** (panel B) illustrates the simulated effects on the incidence rate, assuming the change point occurs after 2013.

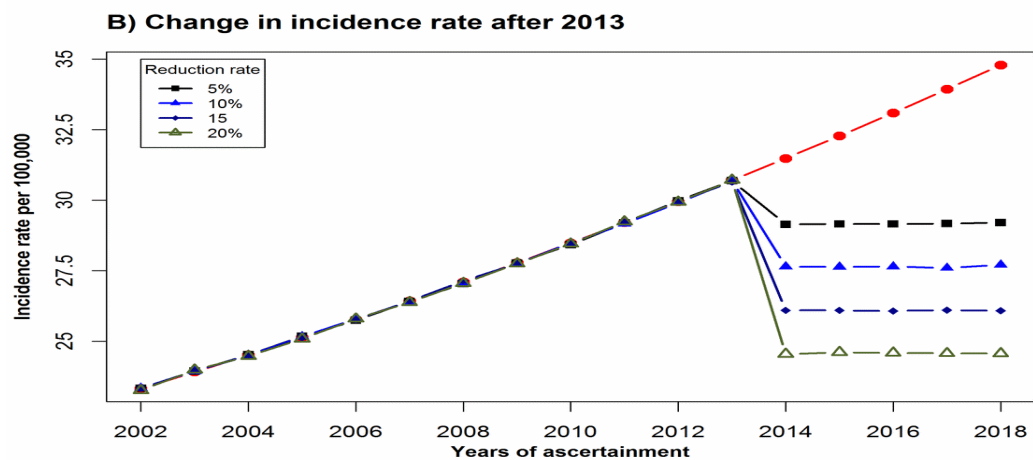*Figure 1.: Observed, fitted, projected and simulated change in incidence rate over time*
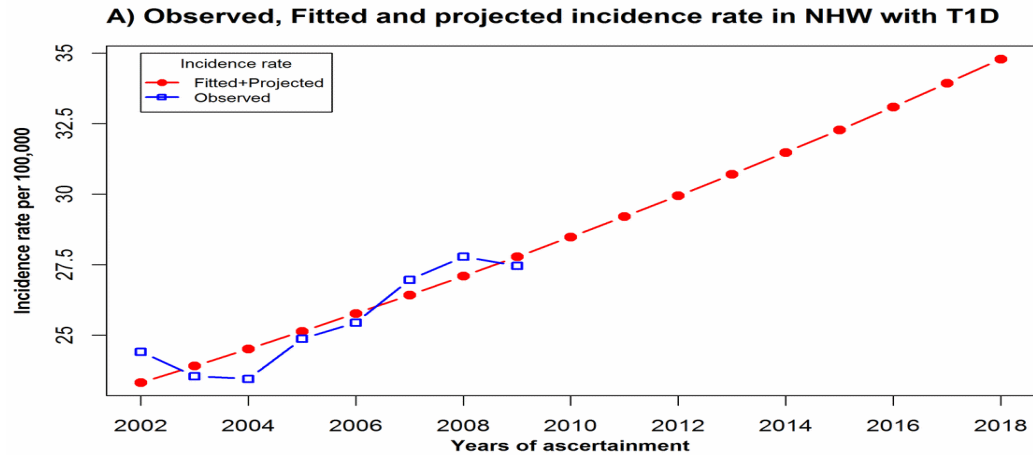
A) Observed, Fitted and projected incidence rate in NHW with T1D

Incidence rate per 100,000

Incidence rate
- Fitted+Projected
- Observed

Years of ascertainment

B) Change in incidence rate after 2013

Incidence rate per 100,000

Reduction rate
- 5%
- 10%
- 15
- 20%

Years of ascertainment

Table 2. shows the estimated power to detect a change point by year of the change point, and the magnitude of the change point, expressed as the percentage of reduction from the last year when an increase in the incidence rate was observed. The power function monotonically decreases with the magnitude of the effect size for the first three years (2012-2014). This paradoxical behavior of the power function for identifying a change point in the distribution of a time series often indicates that nuisance parameters associated with the underlying distribution of the time series could not be estimated (Vogelsang,1999; Crainiceanu et al, 2007).

Table 2.: Power to detect a change point in the incidence rate of type 1 diabetes					
	Year of change point				
Percent reduction	2012	2013	2014	2015	2016
20	0.01	0.07	0.37	0.86	0.99
15	0.05	0.16	0.41	0.77	0.99
10	0.17	0.28	0.50	0.69	0.90
5	0.36	0.41	0.52	0.61	0.69

This is often an indication that the time series is too short to provide reliable estimates of the change point and its standard error. In this case, the dispersion parameter associated with the underlying Poisson distribution used to fit the expected number of non-Hispanic White youths with T1D, and the variance associated with the estimated change point could not be reliably estimated under the null hypothesis. However, the test associated with the estimation of the change point still conserves its good local asymptotic properties, and perform well on longer time series. Here, the monotonic decreasing power function simply suggests that SEARCH will need over 12 years of data before we can begin to consider the change point estimation question. After 2015, the power monotonically increases with effect size. This pattern suggests that we would have 77% power to detect a 15% reduction in the incidence rate after 2015. This power rises to 99% if the change point occurs after 2016. In fact, we would have 90% power to detect a reduction of 10% after that year, and about 70% power to detect a reduction of 5% in the incidence rate. By comparison, the Finnish T1D registry study needed over 30 years of data to retrospectively identify 1988 and 2002 as the years where changes in the incidence rate occurred, with only the change point observed in 1988 being statistically significant (Harjutsalo et al, 2013). The proposed SEARCH analysis would be conducted in the second half of 2019, after the close of the 30-month window needed to register at least 90% of incident 2016 cases.

Identifying a 'leveling off' in the T1D incidence rate in non-Hispanic White youth The power estimation in **Table 2** assumes that the incidence rate is still rising, albeit at a slower rate after a specific year. However, the incidence rates from 2007-2009 in NHW youth with T1D suggest that the incidence rate might have remained constant. Therefore, it is also important to identify the year after which the incidence rate stabilizes. Typical power analysis questions focus on a study's ability to detect a meaningful (pre-specified) effect, instead of the ability to detect non-significant effects. However, equivalence testing provides a framework that allows for power evaluation in this case (Chow et al, 2007; Walker et al, 2011). The null and alternative hypotheses are typically reversed in an equivalence test. For example, for testing the equivalence of two drugs with survival rates P_1 and P_2, the null and alternative hypotheses are stated as follows:

$$H_0 : |P_1 - P_2| > \delta \ vs. \ H_1 : |P_1 - P_2| < \delta \ (\text{Eq. 5})$$ where δ is a tolerance parameter (clinical relevance). In other words, equivalence between P_1 and P_2 is concluded when the null hypothesis is rejected. That is, the difference between P_1 and P_2 does not exceed the range of clinically relevant difference between the survival rates. The tolerance parameter plays the role of the effect size, such that larger sample sizes will be required for smaller delta-values assuming the same difference in the estimated survival rates. Since the tolerance parameter (δ) determine the width of the confidence interval to be constructed around the estimated difference, smaller values of δ result in tighter intervals, which are preferable. The average incidence rate for NHW T1D observed during the 2007-2009 period is 27.5 per 100,000 with a mean absolute difference (MAD) of 0.57. Power estimation assumed $P_1 = 27.5 \times 10^{-5}$; the rate P_2 is then calculated as $P_2 = P_1 + \varepsilon$, where ε varies from 0 to 0.6×10^{-5}. Assuming a significance level α and seven additional incident years of data, the achievable tolerance bounds estimated for 90% and 80% power are given in **Table 3**.

Table 3.: Achievable tolerance with seven additional years of data for an incidence rate of 27.5/100,000 in NHW T1D						
Epsilon	90% Power			80% Power		
	Tolerance	Lower	Upper	Tolerance	Lower	Upper
0	1.62	25.88	29.12	1.44	26.06	28.94
0.1	1.72	25.78	29.22	1.54	25.96	29.04
0.2	1.83	25.67	29.33	1.65	25.85	29.15
0.3	1.93	25.57	29.43	1.75	25.75	29.25
0.4	2.03	25.47	29.53	1.85	25.65	29.35
0.5	2.13	25.37	29.63	1.95	25.55	29.45
0.6	2.24	25.26	29.74	2.06	25.44	29.56
Epsilon: Measure of the variation between estimates (SD)						

For an incidence rate of 27.5/100,000, the achievable tolerance bound is 25.78 - 29.22/100,000, assuming a variation of 0.1 per 100,000 between incidence estimates. That is, if we are willing to consider any value between 25.78/100,000 and 29.22/100,000 to be statistically equivalent to 27.5/100,000, we would have 90% power to detect a leveling-off of the incidence rate of T1D in non-Hispanic White youth with seven more years of data. However, this interval is clearly too wide, which implies that a longer observation period is needed before SEARCH can reliably conclude that the incidence rate of T1D is constant among non-Hispanic White youth. Tolerance bounds for the smaller race/ethnic groups will be even wider. Thus, SEARCH 4 will be unable to detect a leveling off of incidence rates with reasonable confidence. However, it should be noted that, compared to SEARCH, studies who have reported significant changes in their observed incidence rates have had longer follow-up times. For example, the Finland registry report was based on 31 years of data (Harjutsalo et al. 2013).

Estimating Trends in Prevalence
The third prevalence survey of diabetes in youth for SEARCH is scheduled for 2017. Similar to the two previous surveys (2001 and 2009), prevalence will be expressed as the number cases with T1D 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 rates will be assessed by comparing the 2017 estimates to those observed in 2009 and 2001. Chi- square tests will be used to determine if the prevalence of diabetes changed between any two time points, overall and by characteristics of interest. The chi-square test is robust against a large range of possible differences; however, it is not particularly powerful for detecting specific patterns. While numerous patterns of change are possible   (consistently increasing, consistently decreasing, increasing then decreasing, decreasing then increasing, etc.), we are primarily interested in detecting consistently increasing or decreasing changes over time. Poisson regression models will be fit to incorporate results from all three prevalence surveys. Standard errors associated with the estimated change in prevalence rates between any two surveys will be computed using a two-sided skew-corrected inverted score tests for binomial distribution (Gart et al, 1990). Standard errors for the trend in prevalence estimates will be derived from the Poisson regression model. This model will also be used to generate adjusted prevalence rates where adjustment will be made for race/ethnicity, age and sex. The detectable minimum rate of change was estimated using prevalence data collected in 2001 and 2009. Assuming a linear relationship between

prevalence estimates over time, the detectable rate of change was estimated using the approach proposed by Nam[7], which can be seen as a generalization of Cochran-Armitage's trend test for linear trends in proportions. **Table 4.** shows the detectable rate of change by diabetes type and race/ethnicity, assuming a type I error rate of 5% and power levels of 90% and 80%.

Table 4. Detectable Differences in Prevalence by diabetes type and race/ethnicity								
Diabetes type	**Race / Ethnicit**	**Prevalence**			**Δ*2017-2009**	**Δ*2017-2001**	**Detectable rate of change**	
		2001	**2009**	**2017***			**Power =**	**Power = 80%**
Type 1	NHW	1.86	2.55	3.23	26.93	73.70	4.05	3.50
	H	0.96	1.29	1.61	25.14	67.14	8.69	7.51
	NHB	1.29	1.62	1.95	20.49	51.54	11.25	9.73
	API	0.50	0.60	0.69	15.75	37.35	22.45	19.41
	AI	0.30	0.35	0.40	13.83	32.12	43.18	37.32
Type 2	NHW	0.09	0.12	0.16	29.36	83.33	21.18	18.31
	H	0.26	0.53	0.81	51.28	210.39	15.64	13.52
	NHB	0.55	0.75	0.95	26.79	73.20	19.08	16.49
	API	0.21	0.25	0.28	15.07	35.45	40.17	34.72
	AI	0.73	0.82	0.92	11.45	25.84	32.32	27.94
* Prevalence estimated assuming a linear relationship between prevalence rate and time. NHW=non-Hispanic White; NHB=non-Hispanic Black; API=Asian Pacific Islander; AI=American Indian								

The columns denoted $\Delta^*2017\text{-}2009$ and $\Delta^*2017\text{-}2001$ represent the estimated rate of change in the prevalence estimates between 2017 and 2009, and between 2017 and 2001. The estimated prevalence in 2017 was estimated assuming a linear trend. The comparison between those rates and the detectable rate of changes suggests that we are well-powered to detect changes in prevalence for each diabetes type. We are also well- powered to detect linear trends in specific race/ethnic group for each diabetes type. For example, we will have at least 90% power to detect a rate of change of 4.05% in non-Hispanic White (NHW) youth with T1D, and a rate change of 19.08% in non-Hispanic Black (NHB) youths with T2D for the 2017-2009 period. The expected rates of change are 26.93% and 26.79%, respectively. We are even better powered to detect changes in the prevalence rate between 2017 and 2001 since much larger effect sizes are expected.

B.2. Procedures for the Collection of Information

Initial Patient Survey (IPS) (information collection instruments **Attachments 4A1, 4A2a, 4A2b, 4A5a, 4A5b)** contains key data, including the core information described above. In addition, the IPS is used: a) to verify of case eligibility based on residence in year of diagnosis (i.e. confirm that the participant is in the denominator for geographic based clinical sites); and b) is the main source for self-reported race/ethnicity information. Additional information includes:

symptoms at diabetes presentation, potential secondary causes of diabetes, use of insulin, other diabetes medications and any other medications, family structure, usual language spoken, and contact information (for local use only).   This is completed for both incident and prevalent cases.

In-Person Research Visit (IPV) (information collection instruments **Attachments 4A3 and 4A4)** is designed to collect data on relevant characteristics of diabetes type (presence of autoimmunity, genetic susceptibility to autoimmunity, insulin sensitivity, insulin secretion) and data informing the clinical presentation of diabetes. The following will be stored for future analyses (by separate consent): blood, serum, plasma and urine for future genetic and non-genetic analyses. Only diabetes cases incident in 2016 will be eligible to participate in the IPV. An additional sampling approach will be implemented in SEARCH 4, in order to reduce participant burden and maximize study resources, without compromising the statistical power to detect trends in clinical characteristics over time. To maximize the number of minority participants and youth with type 2 diabetes, eligible cases for SEARCH 4 IPV are 100% of minority (non-Caucasian) youth, regardless of age; we will invite to participate in the IPV 100% of Non-Hispanic white youth, aged ≥10 years at diagnosis and 50% of non-Hispanic youth with onset age < 10 years. We will seek a 80% completion of the IPV among eligible youth.

Two additional related activities will be conducted by the clinical sites as part of their Cooperative Agreement responsibilities, but do not directly involve burden to participants:

Collection of Core Variables: A minimum amount of demographic and clinical information is needed for all registered cases in order for the study to be able to calculate population-based rates of diabetes mellitus by age, gender, diabetes type and race/ethnicity for the entire population of cases. This information is also critical in assessing possible response bias to the in-person research visit.

Medical Record Abstraction (MRA) serves the following purposes: a) validation of diabetes diagnosis; b) main source of core demographic and diagnostic information, and c) secondary data source for race/ethnicity information. In SEARCH 2, an additional set of items pertinent to clinical presentation was added to the medical record abstraction effort: weight/height at diagnosis, diabetes ketoacidosis (DKA) at diagnosis and insulin use history. We will continue to collect these data through MRA in SEARCH 4 and will seek 100% completion.

B.3. Methods to Maximize Response Rates and Deal with Nonresponse

To maximize response rate, recruitment materials have been developed that were designed to be culturally appealing to all potential participants. The logo and brochures have been approved by the Steering Committee and distributed to participants and providers by mail as a stand-alone recruitment tool or in conjunction with an introductory letter. The brochures have been made available to potential study candidates in doctors' offices or clinics. They are available in both English and Spanish. The SEARCH sites may translate the brochure and other recruitment materials into additional languages as needed to ensure that all eligible individuals are invited to participate.

A recruitment and retention (R&R) committee was created in SEARCH to ensure complete ascertainment of all cases of diabetes in the eligible study population, maximize the number of registered children, adolescents, and young adults participating in study visits and completing study surveys, and maintain continued contact with participants for long term follow-

up. The SEARCH Coordinating Center works closely with the R&R and project managers to generate recruitment and retention reports by site, incidence year, diabetes type, etc. The R&R provides monthly updates of recruitment and retention efforts to the Steering Committee and advise this committee on appropriate incentives needed to improve the recruitment and retention in specific subgroups.

In order to account for the non-response rate when estimating incidence and prevalence rates, SEARCH utilizes a capture-recapture method. The method begins with the deduplication process. This process, established in SEARCH in 2005, uses date of birth, sex, first two letters of first name and first letter of last name. It was successful in removing almost all duplicates from a database of 10,000 potential SEARCH participants. This deduplication process is crucial for identifying cases found from two different sources within the same SEARCH site, estimating the number of missed cases and the overall capture-recapture rate. This rate is estimated assuming two capture-recapture modes (inpatient and outpatient). A binary outcome (found vs. not found) is defined for each mode leading to a 2x2 contingency table, with the number of participants not found at either source being the unknown variable to be estimated. Once two modes were identified and their duplicates noted, log linear models are fit to the data to estimate the total (unknown) population. The capture-recapture rate can then be used to account for the non-response rate when estimating the incidence and prevalence rates. SEARCH has published a Technical Report detailing these methods (**Attachment 12**).

B.4. Tests of Procedures or Methods to be Undertaken

The procedures and methods of data collection have all been refined previously to minimize burden and improve utility in SEARCH 1, SEARCH 2 and SEARCH 3. Changes to the procedures or methods of data collections being undertaken during the period of data collections being herein requested are detailed in the table below.

Detailed changes to data collection instruments:

Attachment	Form name	Changes made
4A1	Medication Inventory	No changes
4A2a and 4A2b	Incident Case Initial Participant Survey	No changes to the following questions: 1,2,8,9,10
		Deleted question 7 asking participant to report diabetes type
		Deleted question 14 asking participant if s/he was in the military at time of diagnosis
		Deleted question 15 asking participant to report weight
		Deleted question 16 asking participant to report height
		Combined question 22 and 23 into one

		question on health insurance and moved to 12
		Moved the following questions to improve the flow of the questionnaire: 3 moved to 4, 6 moved to 7, 11 moved to 13, 12 moved to 14, 13 moved to 6, 17 moved to 18, 18 moved to 19, 19 moved to 20, 20 moved to 21, 24 moved to 11, 25 moved to 15, 26 moved to 16, 27 moved to 17, 28 moved to 25, 29 moved to 24, 30 moved to 25
4A3	Physical exam	No changes
4A4	Specimen collection	Revised questions 5 and 6 since spot urine is no longer collected. Still ask females if they are pregnant or menstruating at the time of the exam
4A5a and 4A5b	Prevalent Case Initial Participant Survey	New

B.5. Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data

CDC will consult with the SEARCH Coordinating Center, SEARCH clinical sites and CDC partners. The SEARCH Clinical Sites are responsible for the data collection from the participants. The Coordinating Center is responsible collecting the data from the Clinical Sites. Data management and analysis will be performed by the SEARCH Coordinating Center at Wake Forest University. Specific data analysis plans are developed in collaboration with the SEARCH Clinical Sites, the CDC and the Coordinating Center

Lynne Wagneknecht PhD
Wake Forest University School of Medicine– Coordinating Center
Wake Forest University Health Sciences Medical Center Blvd.
Winston-Salem, NC 27157
336-716-7652
lwgnkcht@wfubmc.edu

Ralph D'Agostino, PhD
Wake Forest University School of Medicine– Coordinating Center
Wake Forest University Health Sciences Medical Center Blvd.
Winston-Salem, NC 27157
336-716-9410
rdagosti@wfubmc.edu

Sharon Saydah, PhD
Senior Scientist
CDR USPHS

Centers for Disease Control and Prevention
Division of Diabetes Translation
Chamblee Campus, Blg 107
Atlanta, GA 30341
301-458-4183
ssaydah@cdc.gov

Giuseppina Imperatore, MD PhD
Epidemiology Team Lead
Centers for Disease Control and Prevention
Division of Diabetes Translation
Chamblee Campus, Building 107
Atlanta, GA 30341
770-488-5821
Gai5@cdc.gov