SEARCH FOR DIABETES IN YOUTH STUDY

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

University Park, Williams Bldg.

Atlanta, GA 30341

Telephone: (301) 458-4183

Fax: (301) 458-4025

e-mail: ssaydah@cdc.gov

November 2, 2011

Table of Contents

Attachments

B.1. Respondent Universe and Sampling Methods

Registry Study- *Case Ascertainment Processes utilized by the SEARCH 3 Clinical Sites*

Ongoing Case Ascertainment: SEARCH 3 will use the reporting network of clinics and health care providers that was established in SEARCH phases 1 and 2 as **(Attachment 11)** the primary approach to case-finding for incident cases of diabetes for the period 2010-2014. Additionally, the case ascertainment approach involves existing validated pediatric diabetes databases, hospital and health care plan databases, and other health care organizations.

Case Validation: 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 and 2, 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 255 cases per year.

Cohort Study: Eligibility and Anticipated Sample Size

Individuals registered in the SEARCH cohort with a diagnosis of non-gestational diabetes in incident years 2002-2005, 2006 and 2008 and who completed a SEARCH baseline in-person visit as described in **Attachment 11** in five of the SEARCH clinical centers will be eligible for the new SEARCH Cohort Study once they have had diabetes for at least 5 years. We will determine an algorithm that optimizes duration of disease over the data collection years of the Cohort Study, considering both diagnosis year and the timing of the most recent SEARCH visit during SEARCH 2. We anticipate that the average duration of diabetes will be 8-9 years (range 6-14 yrs diabetes duration), and the average age of participants at the time of the SEARCH Cohort Visit will be about 20 years (range 8-32 yrs old). Our goal is to have 80% of eligible members of the cohort attend the proposed SEARCH Cohort Study visit, for a total of 3145 participants with a Cohort Study visit.

Of the 20% not attending the research clinic visit, an additional 10% will participate by completion of surveys by telephone or internet, for a total of 90% inclusion or 3550 individuals. The estimated number of participants per year (as reflected in the burden table in SS Part A) is based on a 90% response rate of 3550 participants over 5 years of the study resulting in 710 participants per year. Based on proportions of youth in sub-groups of diabetes type and race/ethnicity, the anticipated numbers and characteristics of anticipated participants in the SEARCH Cohort Study data collection are displayed in Table 1.

Table 1. Anticipated Sample Size and Characteristics: SEARCH Cohort Study (2002 - 2006, 2008)

	SC		OH		CO		CA		WA		All	
Type Race/Ethnicity	80%	90%	80%	90%	80%	90%	80%	90%	80%	90%	80%	90%
Type 1												
NHW	405	456	386	434	711	800	98	110	410	461	2010	2261
Other	175	197	58	65	188	211	185	209	94	106	700	788
Type 2												
NHW	38	42	30	34	16	17	7	8	16	18	107	119
Other	106	120	42	47	79	95	83	97	18	23	328	382
ALL	724	815	517	581	999	1124	377	424	541	608	3145	3550

Cohort Study Power Analysis

Based on our calculations from SEARCH 1 and 2 we anticipate that at least 3,288 subjects will participate in the SEARCH Cohort Study in-person visit, and therefore the calculations below use that sample size as the starting value for estimating power and detectable differences.

Incidence Rate Estimation

To estimate power for this component we first had to estimate the proportion of participants who would be free of the condition at their initial in-person visit during SEARCH 1 and 2. Table 2 shows the expected sample sizes available for comparing incidence rates between subgroups under two scenarios: 1) proportion in subgroups are 86% versus 14% (the proportions of T1D and T2D), and 2) proportion in subgroups are 65% versus 35% (the proportions of NHW and all others).

Table 2. Expected number of participants free of outcome at initial visit

	Scenario 1		Scenario 2	
Outcome (% of participants free of condition at initial SEARCH visit)	Subgroup A (86%)	Subgroup B (14%)	Subgroup A (65%)	Subgroup B (35%)
Hypertension (92%)	2601	631	1964	1217
Obese (BMI-z ≥ 95th percentile; 76%)	2149	865	1623	1265
High low density lipids (≥ 100mg/dL; 91%)	1611	955	1217	1180
High ACR (≥ 30; 90%)	2544	668	122	1229
Hypoglycemia in last 6 months (91%)	2573	650	1943	1223
DKA in last 6 months (85%)	2403	751	1815	1251

Using Table 2 we can determine detectable differences for each outcome/group comparison for a variety of plausible scenarios for incidence rates. Table 3 illustrates detectable differences assuming a two group continuity corrected chi-square test for a variety of scenarios with alpha=0.05 (2-sided).

Table 3. Detectable Differences for each outcome

	Scenario 1: Example type 1 vs Type 2		**Scenario 2: Example Non-Hispanic White vs Other**	
Outcome	Incidence for type 1, %	Detectable rate for type 2, % (power)	Incidence rate for NHW, %	Detectable rate for other race/ethnic groups, % (power)
Hypertension	6 **	10 (90)	6 **	9 (86)
	12	17 (88)	12	16 (87)
Obese	5	8 (84)	9	13 (91)
	10	14 (85)	19	24 (89)
High LDL	23	29 (91)	14	19 (90)
	33	39 (85)	24	30 (90)
High ACR	9	13 (83)	5	8 (90)
	19	25 (91)	10	14 (91)
Hypoglycemia in last 6 months	11	16 (91)	23	28 (87)
	21	27 (88)	33	38 (80)
DKA in last 6 months	15	20 (87)	9	13 (92)
	25	31 (88)	19	24 (09)
** **First incidence rate reflects observed incidence rate in SEARCH 1 and 2**				

Based on Table 3, we see, for instance, that there is 84% power to detect a difference between Type 1 and Type 2 participants on their rate of incident obesity if the rate of incident obesity is 5% in the Type 1 group and 8% or higher in the Type 2 group. Likewise, there if the rate of obesity were 9% in the NHW group then there is 91% power to detect a "Other" race/ethnic group rate of 13% or higher. The above calculations should be conservative since when we

adjust for participant level characteristics in our models we should reduce variability and increase precision as we estimate the difference in incidence rates between groups.

Prevalence Estimation

Unlike the incidence rate comparison, all SEARCH Cohort Study participants can be used for the prevalence rate analyses since the incident rate calculations need to remove participants who have the outcome present at visit 1 from the analyses. With this in mind we estimate that there will be 3288 participants available to contribute to prevalence rate estimates. Based on this, Table 4 shows a variety of scenarios for detectable differences comparing groups (i.e., type 1 vs type 2, non-Hispanic white vs others, etc) using a chi-square test to compare groups with alpha=0.05 (2-sided).

Table 4. Scenarios for detectable differences comparing groups

Outcome	**328/2960 (10%/90% split)**		**492/2795 (anticipated type 2/type 1 split)**		**2135/1152 (anticipated NHW vs other race split)**	
	Percent with trait in smaller group					
	10%	20%	10%	20%	10%	20%
Retinopathy	0.16 (83)	0.28 (87)	0.15 (84)	0.26 (80)	0.14 (91)	0.25 (90)
	5%	15%	5%	15%	5%	15%
Neuropathy	0.10 (86)	0.22 (84)	0.09 (86)	0.21 (87)	0.08 (90)	0.19 (82)

Based on Table 4, we see that there is 83% power to detect a difference between Type 1 and Type 2 participants on their prevalence of retinopathy: 10% in the Type 1 group and 16% or higher in the Type 2 group, a realistic potential comparison given early pilot findings of 18% of youth with evidence of DR among the first 38 evaluated.

Longitudinal Models Component

For the purposes of estimating the sample size needed to detect a significant difference with sufficient power, calculations were based on comparing measurements after adjusting for visit 1 data. These calculations need to account for the proportion of the variance in the outcome that is explained by the visit 1 values. Although our full longitudinal models will incorporate all intermediate time points into the final analysis, our power calculation is based on examining

the difference in the outcome of interest adjusting only for the visit 1 assessment of the outcome. Therefore, these power calculations will be conservative, since the additional information provided by the intermediate assessments of outcome measures are not included.

The following formula was used to describe the minimum detectable difference in terms of standard deviations between the participants in groups (i.e., Type 1 versus Type 2). In the formula, r2 is the percent of the variance of the follow-up outcome explained by the visit 1 measurements, $Z_{1-\alpha/2}$ is the value from the standard normal distribution corresponding to the alpha level chosen (1.96, which corresponds to alpha=0.05 [two sided]), $Z_{1-\beta}$ corresponds to the power chosen for the study (80%), σ^2 is the variance of the outcome of interest (i.e. systolic blood pressure), n_1 is the number of participants in the Type 1, k is the ratio of n_1/n_2 (sample size in type 1 and type 2 groups, respectively) and Δ corresponds to the detectable difference in the mean values of the two groups being compared. Using this formula, we examined the detectable differences for several possible r^2 values assuming 80% power and alpha=0.05. From SEARCH 1 and 2, standard deviations for systolic blood pressure, BMI – Z-scores and LDL cholesterol were estimated as 12.7, .85 and 29, respectively. Using these numbers, Table 5 describes the detectable differences if there were 492 participants in the Type 2 group and 2795 in the Type 1 group.

Table 5. Detectable Differences

Detectable differences with 80% power	**Correlation between baseline and follow-up measure**			
Sample size (n1/n2) (2795/492)	0.50	0.65	0.75	0.85
Systolic blood pressure (mmHg)	1.51	1.32	1.15	0.92
BMI (z-score)	0.10 (SD)	0.09 (SD)	0.08 (SD)	0.06 (SD)
LDL (mg/dL)	3.44	3.02	2.63	2.09

As can be seen, if the correlation between the baseline and follow-up measurements is

moderate (.50) then we have 80% power to detect a difference of 1.51 mmHg for the Type 1 versus Type 2 comparison of blood pressure change. As stated above, these estimates should be conservative because when the additional yearly measurements are incorporated into the longitudinal analyses, there will be additional precision which should reduce variability and allow for smaller between group differences to be detected.

B.2. Procedures for the Collection of Information

Registry study

Initial Patient Survey (IPS) contains key data, including the core information described above, and serves to: a) verify of case eligibility (e.g., residence in the year of diagnosis); 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).

In-Person Research Visit (IPV) 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 2012 will be eligible to participate in the IPV. An additional sampling approach will be implemented in SEARCH 3, 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 3 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 70% 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, DKA at diagnosis and insulin use history. We will continue to collect these data through MRA in SEARCH 3 and will seek 100% completion.

Cohort study

Risk Factors for Diabetes Complications (details on measures Appendix B.1)

Laboratory measures:

- Laboratory assays will measure markers of autoimmunity (diabetes autoantibodies, fasting c-peptide, fasting glucose, hemoglobin A1c, lipid profile (total, LDL- and HDL-cholesterol, and triglycerides), apolipoprotein B, LDL particle size and density, adipocytokines (CRP, IL-6, fibrinogen, adiponectin, leptin).

Physical examination measures:

- Standardized anthropometry methods include height, weight, waist circumference
- Systolic and diastolic blood pressure
- Evaluation for acanthosis nigricans
- Assessment of health behaviors(including dietary intake by food frequency, physical activity, alcohol, active smoking and second-hand smoke exposure)
- Retinopathy
- Nephropathy
- Vascular dysfunction (marker of macrovascular disease)
- Markers of neuropathy
- Acute complications (hypoglycemia and diabetic ketoacidosis)

.

Diabetes Treatment & Psychosocial Factors:

- Diabetes treatment regimen and related technologies
- Psycho-social factors
- Socio-cultural factors (household income, , family structure, preferred language, parental and participant-attained education.
- Processes of care and barriers to care

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

Registry study

Methods for case re-ascertainment and capture-recapture methods are utilized to maximize case registration.

Cohort study retention strategies

Retention strategies will include traditional, proven, cohort retention strategies such as: birthday cards, study newsletters, updating contact information annually, utilizing internet-based search systems to locate individuals lost to follow-up, using cell-phone text messaging and e-mail, offering flexible study date appointments including home

visits, offering assistance with transportation, mailing pre-visit instructions, a reminder call prior to the visit, acknowledgement of participation, and participant remunerations that are appropriate for the length and the respondent burden of the proposed study visit. Investigators and study personnel will 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.

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 and SEARCH 2. In addition, the majority of data collection instruments (for example the Centers for Epidemiologic Studies-Depression, Pediatric Quality of Life and the Tanner Stage) have been used and validated in other epidemiologic and clinical studies. There are no new procedures or methods of data collections being undertaken during the period of data collections being herein requested.

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

Ronny Bell, PhD and Ralph D'Agostino, PhD.
Principal Investigators for the Coordinating Center, Responsible for overseeing the data collection from the SEARCH Clinical sites and the data analysis.
Wake Forest University School of Medicine– Coordinating Center
Wake Forest University Health Sciences Medical Center Blvd.
Winston-Salem, NC 27157
336-716-9736
rbell@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
LCDR USPHS
Centers for Disease Control and Prevention
Division of Diabetes Translation
Koger Center, Williams Blg
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
Koger Center, Williams Blg
Atlanta, GA 30341
770-488-5821
Gai5@cdc.gov