Published in final edited form as:
Diabetes Metab Res Rev. 2013 October ; 29(7): 557–567. doi:10.1002/dmrr.2427.

Methods, Quality Control and Specimen Management in an International Multi-Center Investigation of Type 1 Diabetes: TEDDY

Kendra Vehik, PhD[1], **Steven W. Fiske, MA**[1], **Chad A. Logan, MPH**[1], **Daniel Agardh, MD, PhD**[2], **Corrado M. Cilio, MD, PhD**[2], **William Hagopian, MD, PhD**[3], **Olli Simell, MD, PhD**[4], **Merja Roivainen, PhD**[5], **Jin-Xiong She, PhD**[6], **Thomas Briese, PhD**[7], **Sami Oikarinen, MSc**[8], **Heikki Hyoty, MD, PhD**[8,9], **Anette-G. Ziegler, MD**[10], **Marian Rewers, MD, PhD**[11], **Ake Lernmark, PhD**[2], **Beena Akolkar, PhD**[12], **Jeffrey P. Krischer, PhD**[1], **Brant R. Burkhardt, PhD**[1,13], and **the TEDDY Study Group**

[1]Pediatrics Epidemiology Center, Morsani College of Medicine, University of South Florida, Tampa, Florida [2]Lund University, Department of Clinical Sciences, Malmo University Hospital, Malmo, Sweden [3]Pacific Northwest Diabetes Research Institute, Seattle, Washington [4]Department of Pediatrics, University of Turku, Turku, Finland [5]National Institute for Health and Welfare, Helsinki, Finland [6]Medical College of Georgia, Augusta, Georgia [7]Center for Infection and Immunity, Mailman School of Public Health, Columbia University, New York, New York [8]Department of Virology, Medical School, University of Tampere, Tampere, Finland [9]Department of Clinical Microbiology, Centre of Laboratory Medicine, Tampere University Hospital, Tampere, Finland [10]Institute of Diabetes Research, Helmholtz Zentrum München and Forschergruppe Diabetes e.V. Neuherberg, Germany [11]Barbara Davis Center for Childhood Diabetes, University of Colorado Denver, Aurora Colorado [12]Division of Diabetes, Endocrinology and Metabolic Diseases, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland [13]University of South Florida, Department of Cell Biology, Microbiology and Molecular Biology, Tampa, Florida

Abstract

The Environmental Determinants of Diabetes in the Young (TEDDY) is a multi-center, international prospective study (n = 8,677) designed to identify environmental triggers of type 1 diabetes (T1D) in genetically at-risk children from age 3 months until 15 years. The study is conducted through six primary clinical centers located in four countries. As of May 2012, over three million biological samples and 250 million total data points have been collected which will be analyzed to assess autoimmunity status, presence of inflammatory biomarkers, genetic factors, exposure to infectious agents, dietary biomarkers, and other potentially important environmental exposures in relation to autoimmunity and progression to T1D. The vast array and quantity of longitudinal samples collected in the TEDDY study present a series of challenges in terms of

Corresponding Author: Kendra Vehik, PhD, MPH, University of South Florida, Pediatric Epidemiology Center, 3650 Spectrum Blvd., STE 100, Tampa, FL 33612, Telephone: 813.396.2693, Fax: 813.910.5978, kendra.vehik@epi.usf.edu.

Conflict of Interest Statement:
There are no conflicts of interest associated with this manuscript.

quality control procedures and data validity. To address this, pilot studies have been conducted to standardize and enhance both biospecimen collection and sample obtainment in terms of autoantibody collection, stool sample preservation, RNA, biomarker stability, metabolic biomarkers, and T-cell viability. This paper details the procedures utilized to standardize both data harmonization and management when handling a large quantity of longitudinal samples obtained from multiple locations. In addition, we provide a description of the available specimens that serve as an invaluable repository for the elucidation of determinants in T1D focusing on autoantibody concordance and harmonization, transglutaminase autoantibody (tGA), inflammatory biomarkers (T-cells), genetic proficiency testing, RNA lab internal quality control testing, infectious agents (monitoring cross contamination, virus preservation, and nasal swab collection validity), and HbA1c testing.

Keywords

quality control; data integrity; stool sample preservation; RNA; biomarker stability; metabolic biomarkers; T-cell viability

Introduction

The advent of multi-center studies has necessitated more rigorous quality control (QC) methods to minimize inter-laboratory and inter-center variation across highly diverse locations. Extended longitudinal observational study designs add an additional layer of complexity with a high volume of repeated data collected over long periods of time. TEDDY is a unique 15-year longitudinal observational study that spans four countries and 12 clinical sites. This study is setting the benchmark for analyzing many environmental exposures as markers for environmental and lifestyle factors and their association with type 1 diabetes (T1D) [1]. QC efforts at the initiation of the study were necessary to assess the reliability and validity of data/sample collection, shipping, and analyses given the sheer number of biospecimens collected from widespread clinical centers for analysis in multiple core laboratories. Since onset of T1D is believed to be the result of a multi-factorial process of dynamic interaction of both genetic susceptibility and environmental triggers, the objective of the TEDDY study is to identify environmental factors (viruses, bacteria, diet, lifestyle, stress, and chemical exposures) associated with islet autoimmunity and T1D. To accomplish this difficult objective, a long-term prospective observational study was initiated in a high genetic risk cohort. In addition, the study will assess factors associated with celiac disease.

This paper summarizes the technical aspects of the QC methods employed for ensuring accurate HLA screening across clinical centers, autoantibody assay harmonization across core laboratories, stool sample collection, shipment and preservation, RNA quality, biomarker stability, T-cell viability and data quality. The TEDDY Study plans to explore novel data-driven approaches to identify phenotypes linked to T1D susceptibility related to metabolomics, the microbiome, viral metagenomics, epigenetics and gene expression. Biospecimens are stored at the TEDDY Repository managed by Fisher Biosciences. A detailed summary of the available biospecimens collected to assist in the identification of

gene-environment interactions associated with T1D and celiac disease is provided in this paper.

Study Design and Population

TEDDY is a multi-center, multinational epidemiological natural history study initiated by the NIH to identify environmental exposures associated with autoimmunity and T1D onset in children at increased human leukocyte antigen (HLA)-conferred genetic risk for this disease [1]. Newborn screening began in September 2004 and was completed in March 2011. There are 8,677 children enrolled in the 15-year prospective follow-up. There are four countries and twelve centers in Europe (Finland, Germany, and Sweden) and the U.S. (Georgia/Florida, Colorado, and Washington State) and the central TEDDY Data Coordinating Center (DCC) in Tampa, Florida. The TEDDY Study cohort analysis plan incorporates two different statistical designs depending on the nature of the factor under study. For exposures whose values are known on the entire cohort, a prospective design utilizing survival analysis (log rank tests and Cox Proportional Hazards Models), repeated measures analysis of variance, generalized estimating equations and general linear models will be used. For factors related to sample assays, a nested case-control design will be employed. The matched case-control study using conditional logistic regression will involve two interim phases of analysis at 5 and 10 years and the final analysis at the end of the study. The TEDDY Study has entered the first phase of interim analyses with 100 T1D cases and 400 subjects with persistent confirmed autoantibodies with a 1:3 case-control match for the majority of biological samples and a 1:1 match for gene expression, the microbiome and viral metagenomics. Clinic visits began at 3 months of age and continue at a 3-month interval up to the age of 4 years. If subjects seroconvert to persistent autoantibody positivity (GADA, IA-2A or mIAA) they continue on the 3-month interval visit schedule up to age 15 years; otherwise, they switch to a 6-month interval visit schedule (Table 1).

Methods and Quality Control

Genetic Inclusion Screening Quality Control

HLA Screening at Clinical Centers with Confirmatory Testing at a Central Lab

—The TEDDY cohort was enriched for subjects likely to reach disease endpoints by screening for specific HLA genotypes associated with moderate to high future risk of T1D. T1D incidence is relatively low compared to other chronic disease; therefore, TEDDY screened 424,788 children to identify its cohort. A total of 418,367 General Population (GP) infants were screened, of which 20,152 (4.8%) were eligible and 1,437 of the 6,421 screened infants (22.4%) with a First Degree Relative with T1D (FDR) were eligible. The details of this effort have been previously described [2]. The cohort was identified by HLA class II genotyping of newborn screening samples to allow determination of risk based on criteria established with pre-TEDDY data. The HLA typing was completed at five of the six international centers (the Finland Center conducted the typing for the German Center) using either asymmetric polymerase chain reaction and subsequent hybridization of allele-specific probes for HLA-DQA1, DQB1 and DRB1 as described [3] using DELFIA reagents (Perkin-Elmer, Waltham, MA USA) or a dot blot hybridization assay as detailed elsewhere [4]. Both methods used the same study inclusion criteria. There were separate inclusion criteria for the

GP and FDRs of T1D patients. The FDR eligibility includes nine haplogenotypes (*DR3/4*, *DR4/4*, *DR4/8*, *DR3/3*, *DR4/4b*, *DR4/1*, *DR4/13*, *DR4/9*, and *DR3/9*) for broad HLA diversity, whereas the GP eligibility included only the first four haplogenotypes with *DRB1*0403* as an exclusion allele. Errors in screening genotyping could originate from sample mislabeling, true genotyping errors, or rare haplotypes resulting in inferral errors. Central high-resolution confirmation testing was performed on all enrolled subjects and showed that the low-cost and low-resolution genotyping techniques employed at the screening centers yielded an accuracy of 99%. The TEDDY screening strategy demonstrated that different low-cost and low-resolution genotyping methods can result in efficient and accurate identification of a high-risk cohort for follow-up based on the TEDDY HLA inclusion criteria.

Genetic Proficiency Testing—The TEDDY study achieved excellent genotyping accuracy using genetic proficiency testing to ensure high initial and ongoing quality for T1D studies that employ HLA genetic risk assessment [5]. The Centers for Disease Control and Prevention (CDC) conducts both a voluntary quarterly proficiency testing (VQPT) program available to any laboratory and a mandatory annual proficiency testing (PT) challenge for TEDDY laboratories [5]. To mimic and test genotyping samples as those received by TEDDY, CDC sent whole blood and dried blood spots (DBS) samples with a wide range of validated HLA-DR and HLA-DQ genotypes to the five participating laboratories conducting screening tests and the centralized data center. Results were evaluated on the basis of both the reported haplotypes and the HLA genetic risk assessment. In the past six years, the VQPT reported from the 24 panels that 94.7% (857/905) of the relevant HLA-DR or HLA-DQ alleles were correctly identified with 96.4% (241/250) correctly categorized for risk assessment. There was significant improvement seen during the time interval of this program, with correct categorization reaching 100% during the last three years. TEDDY proficiency testing during the past four evaluations has revealed a genotyping accuracy of 99.9% (1153/1154). The different analytical methods used by T1D research centers have all provided accurate (>99%) results for genetic risk assessment. The two complementary CDC PT programs have documented the validity of the various approaches for screening and contributed to overall quality assurance.

Autoantibody Concordance and Harmonization

Autoantibody Harmonization—Participants in TEDDY have autoantibodies measured starting at 3 months of age every 3 months until age 4 years, whereby based on appearance of a single persistent confirmed autoantibody the participant continues on the 3-month interval or if negative transitions to a 6-month interval until the age of 15 years (Table 1). As of May 2012, serum stored in the TEDDY Repository designated for autoantibody testing has been captured on 78% of the cohort adjusted for lost to follow-up (LTF) and withdrawn subjects. In an effort to ensure concordance between the two TEDDY core laboratories that process the autoantibody samples (Barbara Davis Center (BDC), Aurora, Colorado and the University of Bristol Laboratory, Bristol, UK), TEDDY participated in the NIDDK harmonization project in collaboration with the Diabetes Research Institute in Munich, Germany; the details have been published [6]. To evaluate the impact of the harmonized assay protocol on concordance of IA-2A and GADA results, two laboratories retested stored

TEDDY Study sera using the harmonized assays. For IA-2A, using a common threshold of 5 DK units/ml, 549 of 550 control and patient samples were concordantly scored as positive or negative, specificity was greater than 99% with sensitivity 64% in all laboratories. For GADA, using thresholds equivalent to the 97th percentile of 974 control samples in each laboratory, 1051 (97.9%) of 1074 samples were concordant. On the retested TEDDY samples, discordance decreased from 4 to 1.8% for IA-2A (n=604 samples; P=0.02) and from 15.4 to 2.7% for GADA (n=515 samples; P< 0.0001). The precision of the measurement using the harmonized assay was determined in patient samples with values above 2.5 U/ml. The calibration of NIDDK calibrator samples and the median (interquartile range) of the coefficients of variation for IA-2A in the three laboratories (BDC, Bristol, and Munich) were 11.6% (10.6–21.5), 18.8% (15.8–23.9), and 8.7% (6.8–12.9); and for GADA were 10.8% (7.8–17.1), 9.1% (5.2–13.9), and 9% (6–15.3) [6]. The harmonization program for GADA and IA-2A was found to be feasible using large volume working calibrators and similar protocols. It provides a sound approach to ensure consistency in autoantibody measurements. In addition to the harmonized assay, the TEDDY study confirms all autoantibody positive samples and 5% of all negative samples at both laboratories. External quality control is also employed to test concordance by including external control samples identified and aliquoted in Gainesville, Florida and Munich, Germany, embedded in the routine monthly shipments to the two TEDDY core laboratories. These results are compared to external control reference results from the Munich laboratory.

Transglutaminase Autoantibody—Tissue transglutaminase autoantibodies (tGA) are serological markers for celiac disease, a chronic small bowel disease with autoimmune features caused by intolerance to dietary gluten and frequently detected among T1D patients. All children in TEDDY are annually screened for tGA starting at 24 months of age using radiobinding assays (RBA) analyzed in the same two core laboratories as the diabetes autoantibodies. If a child is positive for tGA at 24 months then all of their earlier samples (3 – 21 months) are tested for tGA. A child is classified as having persistent tGA when two consecutive measurements drawn at least three months apart are both positive. Children identified as persistent tGA in TEDDY are referred to a gastroenterologist for management at the clinical discretion of their primary provider. The decision to perform a biopsy is outside the TEDDY study protocol.

At the BDC, the RBA uses anti-IgA agarose to capture IgA-tGA; whereas, in Bristol a mixture of both anti-IgA agarose and protein A sepharose (PAS) is used to assess both IgA-tGA and IgG-tGA. Due to the discrepancy in the detection methods of the two IgA-tGA assays [7], TEDDY uses the BDC laboratory as a screen and the results are based upon the Bristol laboratory's result. Children with a tGA level >0.05 at the BDC and >1.3 units (U) in Bristol are deemed tGA antibody-positive and are re-assayed in a follow-up sample taken after 3 months in samples collected at 24 and 36 months of age and after 6 months from 48 months of age. In order to exclude regional variations in the presence of tGA due to methodological differences between the BDC and Bristol laboratories, persistent tGA-positive sera were analyzed in a small set of samples at both laboratories for confirmatory analysis. In all, 900 TEDDY subjects were included for this exchange analysis and revealed a 93% concordance between the tGA assays, where 54 of discordant children were positive

only in the Bristol and 3 in the BDC assays. Among those 54 children first detected in Bristol but not at the BDC, a total of 27 children reverted to tGA-negative after one or two follow-up samples, 17 remained persistent tGA-positive of which 10 children had subsequent follow-up samples with high tGA levels and were later diagnosed with celiac disease by their physician.

This finding resulted in a protocol modification in order to harmonize the tGA assays in TEDDY and it was decided that Bristol would be used as the reference laboratory in TEDDY for future analyses. The majority of discrepant samples were detected in individuals with low positive tGA levels in Bristol and had a tGA level between the detection level 0.01 and cut-off level 0.05 at the BDC. All sera with tGA levels >0.01 at the BDC are now re-assayed at the laboratory in Bristol for confirmation of persistence. Children found to be persistent positive in Bristol are not re-assayed at the BDC. In discordant cases, sera are further assayed separately for IgG-tGA by the PAS assay in order to distinguish between IgA-tGA and IgG-tGA positivity. Children with initial discordant sera later confirmed by the PAS assay are defined as persistent IgG-tGA positive.

Inflammatory Biomarkers

TEDDY plans to evaluate serum proteins and peripheral blood mononuclear cells (PBMCs) in high-risk children as markers of inflammation and autoimmunity to identify additional biomarkers and time windows of possible therapeutic intervention for T1D.

Cryopreservation of Peripheral Blood Mononuclear Cells (PBMCs)—The PBMC sample collection was initiated in May 2010. PBMCs are isolated from TEDDY children at 3-month intervals using cell preparation tubes (CPT) and a freezing protocol derived from the immune tolerance network (ITN) guidelines [8,9]. To enhance isolated PBMC quality, technicians involved in cryopreservation of PBMCs are required to be annually certified. This certification procedure requires the collection of whole blood in two CPT from three separate non-TEDDY donors and the preservation of PBMCs in three separate vials frozen according to the TEDDY Manual of Operations (MOO). The samples are maintained at −80°C for two days before shipment to the central laboratory (Benaroya Research Institute, Clinical Core Lab, Seattle, USA) for quality control. Cells are evaluated for viability and recovery by certified technicians. The samples with documented sample volume and count information are shipped to the Benaroya Research Institute for centralized analysis. A passing score is determined by both cell recovery and viability with results reported by the DCC.

To optimize the recovery of PBMCs from CPT, a pilot study was conducted to define the best method for cell isolation and to minimize cell loss following centrifugation. Blood samples from 10 healthy donors were collected in three separate CPT and after centrifugation PBMCs were collected in three different ways: (1) removing the PMBCs layer without mixing the cells with the autologous plasma, (2) removing the cells after re-suspension in the autologous plasma by gentle pipetting, and (3) collecting cells as outlined above but the tubes were further washed twice with 2 ml of RPMI medium [9] to collect residual cells. Cell count and viability were determined by trypan blue exclusion assay with

hemocytometer. The results demonstrated that a considerable amount of PBMCs remained in the CPT post-centrifugation and PBMC recovery was increased from 14% to 88% with inclusion of two additional washes of the CPT sample. Therefore, this improved protocol was added to the TEDDY MOO. As of May 2012, 45% of enrolled subjects have at least one preserved PBMC specimen.

RNA Lab Internal QC Testing Results

The TEDDY RNA Laboratory at Jinfiniti Biosciences in Augusta, GA (www.jinfiniti.com) has developed a high-throughput (96-well format) extraction protocol using magnetic (MagMax) beads technology (proprietary technology). Each plate of extraction contains 92 TEDDY samples, two positive control samples prepared by the TEDDY RNA Laboratory and two negative controls (PBS buffer). To assess the quality and quantity of the isolated RNA samples from whole blood and to monitor potential degradation over time, the lab has adopted a series of quality control (QC) procedures. For each extraction plate, all 96 RNA samples are analyzed using a spectrophotometer (NanoDrop) at 260nm and 280nm to document RNA concentration and quality in each sample. These concentrations are plotted in the TEDDY RNA Laboratory Information Management System (LIMS) and negative control samples are used to assess potential cross-contamination during the entire process. The concentrations for the positive controls are compared to historical data for samples obtained from the same positive control individual and these data are used to assess the overall quality of the extraction run. Failure of positive controls as well as TEDDY samples indicates technical problems associated with the run. A 260nm/280nm ratio is calculated for each sample to assess the quality of the RNA samples. Ratios above 2.3 or below 1.6 are indications of poor quality and are flagged in the TEDDY RNA LIMS. The frequency and distribution of poor quality samples are calculated for every 12 consecutive plates (1104 samples) to assess technical and biological variation over time. Two TEDDY samples (one low and one high concentration) are also randomly selected from each plate to be analyzed by the Agilent 2100 BioAnalyzer. Samples from every 12 plates are batched for a BioAnalyzer run. Once a year, 24 TEDDY RNA samples (11 fresh samples, 11 samples stored at −80°C for 6–12 months, and two of the samples from pools of RNA from positive controls that have been created) are analyzed by real-time RT-PCR to assess the quality of RNA using several different genes (CDK2, GAPDH, HPRT1). Annually, 12 TEDDY RNA samples are analyzed by microarray to assess their quality. On average the total RNA isolated is 10.8μg with an average 260nm/280nm ratio of 2.07. No significant degradation of the RNA has been observed over three years, and longer term stability is being monitored.

The TEDDY Study has collected 2.5 ml of peripheral blood to extract total RNA from children enrolled in the study based on the visit schedule outlined above and in Table 1. As of May 2012, 73% of the one or more persistent autoantibody positive subjects have mRNA samples available. These samples have been allotted for gene expression studies.

Infectious Agents

Many prior studies have linked viruses to both autoimmunity and T1D [10–12]. Thus, it is of great importance that the capture, processing and detection of infectious agents provide an optimum measurement and identify a valid exposure.

Preservation of Enteroviruses in Stool Samples during Shipment—Stool samples are collected from all TEDDY children starting at 3 months of age for virological and other microbiological analyses. Samples are taken at home by the parents and sent to the TEDDY Repository by express mail (U.S. centers) or to the local repository by regular mail (European centers) where exposure to ambient temperatures for 1–3 days is possible. The effect of exposure to varying temperatures on enterovirus detection was evaluated by spiking enterovirus negative stool samples or sterile water samples with infective enterovirus (Coxsackievirus B3) and then exposing them to different temperatures for 8–72 hours. The varying temperatures evaluated were: 4.0°C/39.2°F, 25.0°C/77.0°F, 35.0°C/95.0°F, 43.0°C/109.4°F, 56°C/133°F and 65.0°C/149.0°F, at 8, 24, 48 or 72 hours.

The preservation of enterovirus RNA was first analyzed in spiked water samples by a sensitive semi-quantitative RT-PCR (5). The amount of viral RNA remained stable at temperatures ≤35°C/95°F for all tested time points. However, exposure to temperatures ≥43°C/109°F decreased the level of viral RNA detected, especially when exposed for 24 hours or longer. The highest temperatures (56°C/133°F and 65°C/149.0°F) rapidly decreased the amount of viral RNA even after the shortest (8 hour) exposure. The results were similar for the spiked stool samples.

The preservation of viral infectivity was analyzed by plaque assay (the number of plaque forming units in green monkey kidney (GMK) cells). Exposure to temperatures ≥56°C/133°F led to a rapid and complete loss of infectivity at 8 hours exposure. Exposure to 43.0°C/109.4°F for durations <72 hours reduced infectivity by 50–80% and durations ≥72 hours led to a complete loss of infectivity. Exposure to temperatures ≤25.0°C/77.0°F did not reduce infectivity even after 72 hours.

Overall, exposure to temperatures exceeding 35°C/95°F decreased both viral RNA and viral infectivity, and higher temperatures (56°C/133°F or higher) inhibited sensitive detection. Conversely, viral RNA and infectivity remained stable at lower temperatures. These findings resulted in the use of ice-gel stool sample shipments during the summer months to eliminate deleterious temperature peaks. The maximal temperature to which the samples were exposed during regular ice-gel packed shipments was monitored using commercially available CelsiStrip® temperature recording labels. Change in the color indicated the maximal temperature of exposure. These stickers were placed inside the stool sample mailing box during summer months in 1,467 shipments; the majority of samples (99%) remained at a temperature <40°C/105°F.

Monitoring Cross-contamination in TEDDY Blood and Stool Samples—This quality control protocol regularly monitors possible contamination of virus-negative samples by virus-positive samples (cross-contamination). This can happen when many samples are processed at the same time in the same location (e.g., when sample aliquots are made at TEDDY sites). Per protocol sterile phosphate buffered saline (PBS) samples are processed on a monthly basis in the same location as clinical samples. Possible contamination is then monitored by detecting enterovirus and rotavirus in these samples using sensitive RT-PCR. These viruses are common in children and are therefore optimal markers for contamination risk. PBS samples are processed in all sites which process blood, serum, plasma or stool

samples in open tubes (e.g., dividing samples into smaller aliquots by pipettes or processing stool samples from diapers). PBS is first poured from a 500 ml bottle into a 50 ml tube in the same table or laminar flow where the samples are processed, and divided further into ten aliquots using the same procedures and equipment as for clinical samples. Samples are then shipped to the TEDDY Repository and stored at −80°C. Samples are analyzed regularly for the presence of rotavirus and enterovirus at the University of Tampere, Medical School (Tampere, Finland) using established RT-PCR methods validated in external quality control rounds (QCMD; http://www.qcmd.org or equivalent). If needed, the samples can also be screened for other infectious agents. To date, 492 samples have been tested, and all have been virus-negative.

Viral and Pathogen Detection Using Nasal Swabs—This pilot study evaluated the validity of identifying respiratory viruses and other pathogens using nasal swabs. Previous studies have found that nasal swabs are as effective as throat swabs for collecting samples from subjects with active upper respiratory infection; however, these methods have not been applied in regular sampling of asymptomatic subjects. Nasal swab sampling may be a non-invasive way to survey agents frequently causing infection and fever during childhood and the prevalence in the community during defined time periods.

Samples were collected using Copan nasal swabs from 50 healthy subjects (age 1.5 – 14 years, 23 March – 1 April 2009) and 40 subjects with active upper respiratory tract infection (age 0.5 years – adults, 28 May – 5 November 2009). All subjects were from Turku, Finland. Laboratory analyses were performed using virus specific PCR assays (RT-PCR) and multiple pathogen detection (multiplex PCR). For purposes of comparison, RT-PCR was performed at two independent sites in Finland (National Institute of Health and Welfare in Finland and University of Tampere, Finland). Multiplex PCR was performed at a third independent laboratory (Columbia University, New York) [13]. Results from all laboratories confirmed that respiratory pathogens can be detected using nasal swab procedures. The most frequently identified pathogens were human rhinovirus (HRV), *H. influenzae*, and *S. pneumoniae*. Viral pathogens were more frequently reported in subjects with active upper respiratory infection than in healthy individuals (HRV: 26/9 vs. *H. influenzae*: 15/2, *S. pneumoniae*: 2/6). Test accuracy was evaluated by false positive rates (FPR) and false negative rates (FNR). Using single pathogen detection in at least one lab to test subjects for clinical infections resulted in a FPR as high as 46/100 and a FNR of at least 22/100. However, estimating the frequency of asymptomatic and infectious subjects based on detection of multiple pathogens resulted in confirmed pathogen detection in 30/100 subjects with a low FPR of 6/100. Comparison of results across all three laboratories showed a high level of concordance with respect to HRV results. Confirmed positive HRV was detected in 65/100 symptomatic subjects with a FPR of 14/100. The virus specific PCR assay had an overall concordance rate of 98/100 for HRV. This assessment confirms that viral and bacterial pathogens can be detected using nasal swabs in both symptomatic and asymptomatic subjects.

Beginning at 9 months of age a minimally invasive nasal swab sample is collected from each TEDDY subject and continues to be collected on the visit schedule as outlined under Study

Design and Population. Nasal swab collection began in December 2008 and 67% of the enrolled cohort have available samples for analysis.

HbA1c Methods and Quality Control

The longitudinal design of TEDDY provides an ideal cohort to assess metabolic biomarkers and their utility in diagnosis and prediction. HbA1c values reflect a 90-day moving average of blood glucose concentrations, weighted more heavily towards the last 30-days. The TEDDY HbA1c samples are processed at the Diagnostic Diabetes Laboratory (Columbia, Missouri). The instrument Tosoh G7 HPLC analyzer, used for the measurement of HbA1c, is calibrated using two whole blood calibrators - PLC and PHC Lot #4 (PLC4 and PHC4) using the target values that are accepted based on the National Glycohemoglobin Standardization Program (NGSP) network reference. The same set of calibrators is used for the TEDDY Study. As newer instruments are introduced, the instruments will be certified as secondary reference laboratory methods for the NGSP before it can be used for routine analysis of clinical specimens. In addition, the laboratory participates in the College of American Pathologists GH2 survey twice a year as well as in the International Federation of Clinical Chemistry HbA1c monitoring program. Two samples are analyzed each month. Furthermore, the Laboratory also participates in NGSP Network Monthly Monitoring every month using the Ultra 2, Tosoh G7 and G8 HPLC methods which are NGSP Secondary Reference Laboratory (SRL) methods. Monthly surveys are administered by the Central Primary Reference Laboratory (CPRL) to monitor performance of the network laboratories. Each month, the CPRL ships 10 specimens (fresh or fresh/frozen whole blood) over the desired clinical range (4–10% HbA1c) to each network laboratory. The CPRL, PRLs and SRLs analyze the specimens in two separate runs on two separate days. All data are sent to the NGSP Administrative Core for analysis. To maintain certification, a network laboratory must fulfill the following bias and precision requirements: 1) the mean of the differences (n=10) between the network laboratory and the CPRL must not exceed 0.35% HbA1c and 2) the estimate of the standard deviation of the difference in sample replicates must not exceed 0.229 (99th percentile of the sampling distribution around a target SD of 0.15). SRL results must also fall within a defined acceptance ellipse based on the slope and intercept of the differences between the individual SRLs results and the medians of all SRLs.

The TEDDY Study collects a whole blood sample for HbA1c testing on children who are autoantibody positive starting as early as their 12-month visit and then every visit thereafter. TEDDY began HbA1c collection in April 2009. Fifty-five percent of antibody-positive children have samples available.

TEDDY Sample Management

As described in Figure 1, the oversight and coordination of sample management, shipment, and accountability is directed by the DCC. Prior to collection of various patient samples, the DCC directs the shipment of appropriate storage containers to the various patient collection sites. All samples are aliquoted into dedicated, barcoded (Symbol LS 2208), and color-coded cryovials as determined by the type of particular analyses to be performed by selected laboratories. Each cryovial has a distinct and unduplicated barcode with a color-coded top that allows for identification, tracking, and differentiation among the thousands of samples

collected. In addition, to promote the longevity of the samples obtained, preservatives are added at the collection stage to enhance long term stability of the particular analyte. For example, plasma samples assayed for ascorbic acid analysis contain 0.2 ml of 5% trichloroacetic acid and 200 mg disodium EDTA/L. To preserve fatty acid analysis, collected erythrocytes contain 2 ml of 2-propanol with 50 mg/L of butylated hydroxytoluene. Following collection of patient samples, data are transmitted to DCC containing all relevant information regarding the sample collected including barcode, sample type, sample volume, cryovial appearance, date of collection, and patient information. Following transmission of data to DCC, the information is evaluated for potential errors and subsequently stored. The patient samples are then shipped frozen to the TEDDY Repository and immediately stored at −80°C. Based on the format of the experimental procedure, the DCC generates a list of required samples (patient and QC) for necessary analyses and transmits that information to the TEDDY Repository. The TEDDY Repository then removes those samples from storage, confirms the barcode and content and then ships to the selected laboratory for experimental analysis. All data files including initial raw data and modified datasets generated by the selected laboratories are transmitted and stored at DCC.

The DCC has compiled samples strictly for quality control purposes that have been incorporated into the overall sample management (Figure 1) to measure and determine inter- and intra-assay variability from the data received from the various selected laboratories. These QC samples appear identical and similar in biological nature to actual patient samples. Therefore, these samples (plasma, RBCs, stool, RNA, and infectious agents) are processed in the same manner and are stored in sample-matched barcoded vials. Also, pseudo patient IDs and site visits are incorporated for each QC sample to mimic actual patient samples and prevent the laboratory from distinguishing QC samples. QC samples are then integrated into DCC established experimental design and prepared by the DCC QC laboratory and shipped to the TEDDY Repository for storage at −80°C. QC samples are then integrated into the patient sample sets sent to the core laboratories for analysis as previously described (Figure 1). Data generated from the analysis of these QC samples are sent to the DCC as described for patient samples.

Data Collection and Analysis

With over 250 million data points collected thus far, the QC methods employed throughout the collection, entry, and management of TEDDY data are of paramount importance to the overall integrity of the study.

Data are extracted by trained staff members during scheduled visits and entered either directly via standard forms (web forms or through teleforms) which are scanned and transmitted electronically. Front-end constraints are employed in the web application to prevent the entry of invalid data; for example, certain fields are required, only valid dates may be entered, and ages must fall within a predefined range. All TEDDY data are stored in fully managed Oracle databases at the DCC and archived regularly, both on- and off-site, to ensure data security. In addition to the front-end constraints applied at the time of entry, a unique automated QC system has been developed for the TEDDY project.

The TEDDY Error Reporting and Verification System (ERVS) consists of a set of programs that conduct automated QC on TEDDY data, a specialized web application for reporting and resolving errors, an integrated database for storing error data, and a set of programs that generate reports for monitoring data cleaning efforts. Manifold error checks are conducted as a part of this automated QC processing. These include, for example, comparing dates against the visit date, the subject's birth date, and the current date; checking answers to responses on related questions for conflict; identifying outliers through a variety of techniques particular to the circumstance; and verifying all codes using existing code databases. Any detected errors are stored in the ERVS database and are automatically available through the web application. Staff members at the clinical centers then use the web application, along with the original paper records stored at their site, to correct or verify all reported errors. The process is monitored and directed throughout using extensive ERVS summary reports that detail the number of errors identified, pending, corrected, and verified across clinical centers, forms, and even individual questions.

The ERVS allows for timely and directed data cleaning while reducing the burden on both the clinical centers and the DCC. The process is automated and standardized and has resulted in a significant acceleration in manuscript development. Real-time QC processing enables the immediate identification of data collection or entry issues. This assists in maintaining the interpretation of the MOO and allows for targeted retraining of staff; both essential to maintaining data integrity in a massive, multinational, longitudinal study. Critically, the ERVS has resulted in the correction or verification of 117,545 data entry errors; a relatively small (0.28%), but nonetheless significant, portion of the total number of data points checked.

TEDDY analyses data are further subjected to external QC as a part of the data sharing process. Raw manuscript datasets are submitted to the NIDDK Data Repository where they undergo a "Dataset Integrity Check (DSIC)." The DSIC is conducted by NIDDK affiliated statisticians who independently replicate the published results using the raw datasets provided.

Discussion

The TEDDY project represents a unique and massive undertaking to unravel critical clues toward the elucidation of the causative mechanisms of T1D. Due to the multi-factorial nature of this disease, it has become necessary to collect a multitude of specimens longitudinally from numerous centers. This has posed serious challenges in quality control. The solution entails multiple quality control steps for each data point and specimen collection with continual evaluation of archived samples. In doing so, TEDDY has built a generous repository that can serve as the basis for a multitude of studies understanding the determinants responsible for T1D (Table 2). The robust quality control effort should greatly enhance the value of the samples and data collected by TEDDY as a key resource to investigators proposing innovative hypotheses concerning candidate environmental and genetic factors.

Acknowledgments

The TEDDY Study Group (See online appendix). Funded by U01 DK63829, U01 DK63861, U01 DK63821, U01 DK63865, U01 DK63863, U01 DK63836, U01 DK63790, UC4 DK63829, UC4 DK63861, UC4 DK63821, UC4 DK63865, UC4 DK63863, UC4 DK63836, UC4 DK63790, and UC4DK095300 and Contract No. HHSN267200700014C from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Institute of Child Health and Human Development (NICHD), National Institute of Environmental Health Sciences (NIEHS), Juvenile Diabetes Research Foundation (JDRF), and Centers for Disease Control and Prevention (CDC).

Reference List

1. The TEDDY Study Group. The Environmental Determinants of Diabetes in the Young (TEDDY) Study. Ann N Y Acad Sci. 2008; 1150:1–13.
2. Hagopian WA, Erlich H, Lernmark A, et al. The Environmental Determinants of Diabetes in the Young (TEDDY): genetic criteria and international diabetes risk screening of 421 000 infants. Pediatr Diabetes. 2011
3. Kiviniemi M, Hermann R, Nurmi J, et al. A high-throughput population screening system for the estimation of genetic risk for type 1 diabetes: an application for the TEDDY (the Environmental Determinants of Diabetes in the Young) study. Diabetes Technol Ther. 2007; 9:460–472. [PubMed: 17931054]
4. Erlich H, Valdes AM, Noble J, et al. HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk: analysis of the type 1 diabetes genetics consortium families. Diabetes. 2008; 57:1084–1092. [PubMed: 18252895]
5. Dantonio P, Meredith-Molloy N, Hagopian WA, et al. Proficiency testing of human leukocyte antigen-DR and human leukocyte antigen-DQ genetic risk assessment for type 1 diabetes using dried blood spots. J Diabetes Sci Technol. 2010; 4:929–941. [PubMed: 20663459]
6. Bonifacio E, Yu L, Williams AK, et al. Harmonization of glutamic acid decarboxylase and islet antigen-2 autoantibody assays for national institute of diabetes and digestive and kidney diseases consortia. J Clin Endocrinol Metab. 2010; 95:3360–3367. [PubMed: 20444913]
7. Li M, Yu L, Tiberti C, et al. A report on the International Transglutaminase Autoantibody Workshop for Celiac Disease. Am J Gastroenterol. 2009; 104:154–163. [PubMed: 19098864]
8. Immune Tolerance Network. CPT without Plasma Collection. 2006. Isolation and Cryopreservation of PBMC. [v. 009]Ref Type: Online Source
9. Mallone R, Mannering SI, Brooks-Worrell BM, et al. Isolation and preservation of peripheral blood mononuclear cells for analysis of islet antigen-reactive T cell responses: position statement of the T-Cell Workshop Committee of the Immunology of Diabetes Society. Clin Exp Immunol. 2011; 163:33–49. [PubMed: 20939860]
10. Tauriainen S, Oikarinen S, Oikarinen M, Hyoty H. Enteroviruses in the pathogenesis of type 1 diabetes. Semin Immunopathol. 2011; 33:45–55. [PubMed: 20424841]
11. Oikarinen S, Martiskainen M, Tauriainen S, et al. Enterovirus RNA in blood is linked to the development of type 1 diabetes. Diabetes. 2011; 60:276–279. [PubMed: 20943747]
12. Stene LC, Oikarinen S, Hyoty H, et al. Enterovirus infection and progression from islet autoimmunity to type 1 diabetes: the Diabetes and Autoimmunity Study in the Young (DAISY). Diabetes. 2010; 59:3174–3180. [PubMed: 20858685]
13. Briese T, Palacios G, Kokoris M, et al. Diagnostic system for rapid and sensitive differential detection of pathogens. Emerg Infect Dis. 2005; 11:310–313. [PubMed: 15752453]

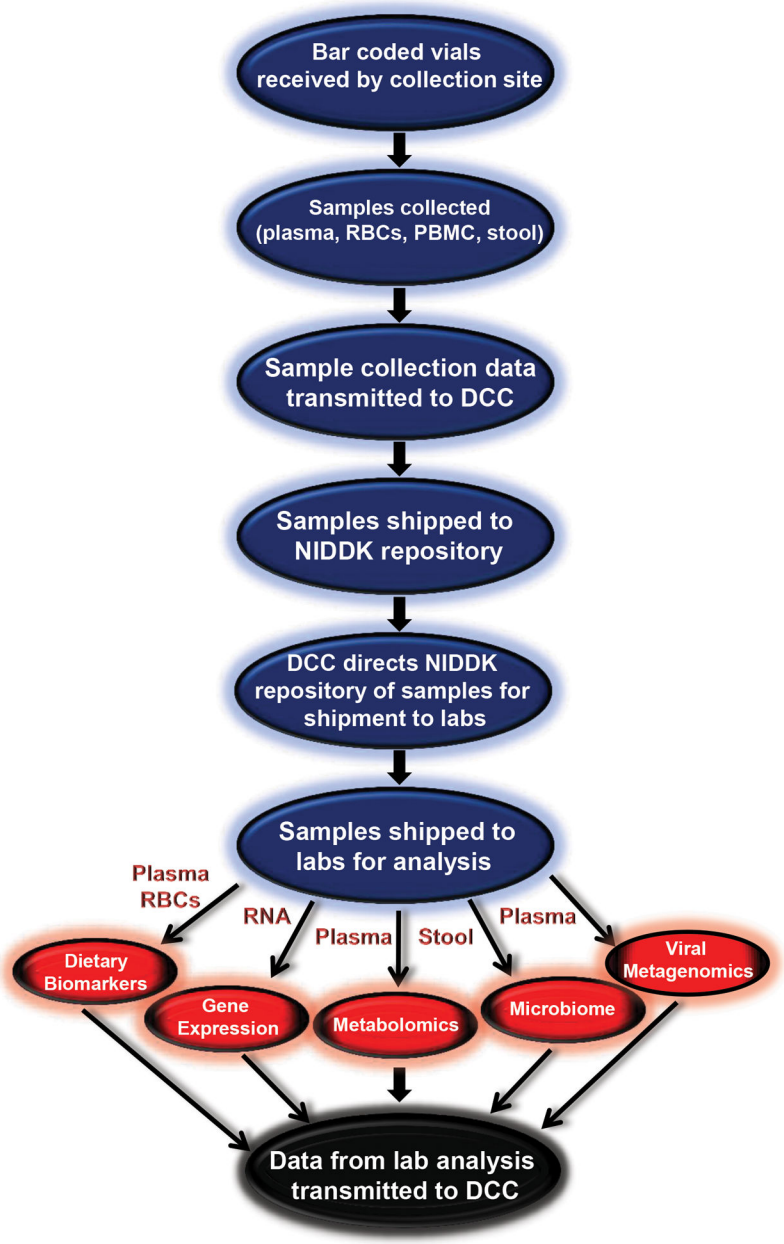

Figure 1. Outline of patient sample management for TEDDY

An overall flow chart details the organizational procedure for handling of various patient samples in terms of collection, shipment, analysis, and record keeping. DCC, Diabetes Coordinating Center

Table 1

Blood Sampling Frequency and Volumes (all volumes are shown in milliliters)

		Age in Months for Time of Collection																		
Analyte	**Sample Type**	**Screening: Birth**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**15**	**18**	**21**	**24**	**24–48 mo Every 3 mo Tests**	**>48 mo Every 6 mo Tests**	**>24 mo Yearly Tests**	**4 Years**
Cord Blood		X																		
Venous/capillary Blood (ml)	Non-fasting	X*	X			X			X			X	X	X	X	X	X#	X#		
Additional HLA[2]	Whole blood					0.5^			1^			1^	1^							
Autoantibodies[1]	Serum		0.2			0.2			0.2			0.2	0.2	0.4	0.4	0.42	0.4	0.4	0.02**	
Serum cytokines/inflammation marker[1]	Serum		0.1			0.1			0.1			0.1	0.1	0.1	0.1	0.1	0.1	0.1		
Entero- & rotavirus																				
PCR	Plasma		0.3			0.3			0.3			0.3	0.3	0.3	0.3	0.3	0.3	0.3		
Antibodies[1]	Plasma		0.1			0.1			0.1			0.1	0.1	0.1	0.1	0.1	0.1	0.1		
Additional infectious agents[1]	Plasma		0.4			0.4			0.4			0.4	0.4	0.4	0.4	0.4	0.4	0.4		
Vitamin D[1]	Plasma		0.05			0.05			0.05			0.05				0.05			0.05	
Alpha-tocopherol, gamma-tocopherol[1]	Plasma					0.07						0.07				0.07			0.07	
Carotenoids[1]	Plasma					0.06						0.06				0.06			0.06	
Ascorbic acid[1]	Plasma					0.05						0.05				0.05			0.05	
RBC Membrane Fatty Acid[1]	RBC		0.5			0.5						0.5				0.5			0.5	
MRNA (Paxgene tube)	Whole blood		2.5			2.5			2.5			2.5	2.5	2.5	2.5	2.5	2.5	2.5		
Non-HLA genotyping[2]	Whole blood																			5

		Age in Months for Time of Collection																		
Analyte	**Sample Type**	**Screening: Birth**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**15**	**18**	**21**	**24**	**24–48 mo Every 3 mo Tests**	**>48 mo Every 6 mo Tests**	**>24 mo Yearly Tests**	**4 Years**
	Total serum	-	1.0			1.0			1.5			3.0	3.0	3.3	3.3	3.3	3.3	3.3		
	Total plasma	-	1.0			2.0			2.5			4.0	4.0	4.0	6.0	6.0	6.0	6.0		
	Total whole blood	-	2.5			2.5			3.5			2.5	2.5	2.5	2.5	2.5	2.5	2.5		
	PBMC[1]		X			X			X			X	X	X	X	X	X	X		
	Total RBC	-	1.0			2.0			4.0			4.0	4.0	4.0	4.0	4.0	4.0	4.0		
	Serum tube		2.0			2.0			3.0			6.0	6.0	6.0	6.0	6.0	6.0	6.0		
	Plasma tube		2.0			4.0			5.0			8.0	8.0	8.0	12	12	12	12		
	ABI tube		2.5			2.5			2.5			2.5	2.5	2.5	2.5	2.5	2.5	2.5		
	HLA confirmation					0.5^			1^			1^	1^							
	Total Blood Volume	-	**6.5**			**8.5 or 9.0^**			**10.5 or 11.5^**			**16.5 or 17.5^**	**16.5 or 17.5^**	**16.5**	**20.5**	**20.5**	**20.5**	**20.5**		
Blood Glucose		**At every visit once subject tests positive for any autoantibody**																		
OGTT (fasting)		**Every six months once subject tests positive for two autoantibodies, regardless of autoantibody positivity confirmation or persistence, at any pervious visit and is three years of age or older**																		
HbA1c [2]		**0.25 mL sample taken at every visit from children who are positive at the 9 month visit or later for at least one autoantibody (regardless of autoantibody positivity confirmation or persistence)**																		

*If cord blood is not available for HLA typing then capillary blood should be drawn.

**Additional 0.02 ml for tissue transglutaminase antibodies measurement added to the islet antibody sample sent to the Autoantibody Reference Lab.

#Children four years of age and older who have been deemed persistent autoantibody positive will remain on the three month visit schedule.

^Only one HLA confirmation sample is needed from the earliest visit with a full volume blood draw.

[1]Cryovial

[2]EDTA tube

Table 2

Detailed Summary of Quality Control and Availability of Biospecimens (sample availability current as of May 2012)

Biospecimen	Quality Control	Final Product	Sample availability in Cohort or Nested Case-Control
Peripheral blood mononuclear cells (PBMC)	Blood collection from TEDDY Manual of Operations with certified technicians followed by cell viability assay	Recovery of cells maintenance of good viability, functionality control (mean recovery >60% and >90% viability) – 5–10 × 10^6 PBMC / ml / 2–5 vials per subject	45% cohort
Human Leukocyte Antigen	Proficiency tested whole blood and dried blood spots (DBS) samples with a wide range of validated HLA-DR and HLA-DQ genotypes	99% accuracy in identifying relevant HLA-DR or HLA-DQ alleles	100% cohort
Diabetes Autoantibodies	Concordance of GADA and IA-2A retested by two laboratories	Concordance with harmonized assay for GADA is 98.2% and 97.3% for IA-2A. The harmonized assay is used for TEDDY protocol.	78% cohort
Trans-glutaminase Autoantibodies (tGA)	Inter-laboratory tGA assay concordance	93% concordance between two tGA assays	100% starting at 24 months of age
Infectious Agents - Enterovirus	Preservation of enterovirus after exposure to varying temperatures and time periods during stool sample transport using reverse transcription polymerase chain reaction (RT-PCR) in spiked water samples	At +4°C/39°F and +25°C/70°F both RNA and infectivity preserved for 72 hours. Samples per protocol will not be exposed to temperatures of 43°C/110°F or higher. Ice-gel is now included in the TEDDY stool sample shipments during the hot summer months to eliminate deleterious temperatures peaks.	81% Enterovirus Antibodies and PCR – case-control
Infectious Agents – Nasal Swabs	False positive and false negative rates of virus/ pathogen detection; frequency of asymptomatic and infectious subjects with multiple viruses or pathogens; and concordance of sample results across three labs	Respiratory pathogens can be frequently detected in nasal swabs using both multiplex and pathogen specific molecular methods.	67% cohort
mRNA	Absorbance 260/280 followed by Agilent 2100 Bioanalyzer	High quality 10.6 ug/ blood sample stored for gene expression studies	73% case-control
Glycated haemoglobin (HbA1c)	Certification maintained through the National Glycohemoglobin Standardization Program Administrative Core	The mean of the differences in samples (n=10) between the network laboratory and the Central Primary Reference Laboratory must not exceed 0.35% HbA1c and the estimate of the standard deviation of the difference in sample replicates must not exceed 0.229 (99th percentile of the sampling distribution around a target standard deviation of 0.15). coefficient of variation<1.3%.	55% autoantibody positive subjects in TEDDY cohort