

Sample and Data Sharing Policy

One of the main goals of “The Environmental Determinants of Diabetes in the Young” (TEDDY) is to establish a central repository of data and biologic samples for subsequent hypothesis-based research. The repository will facilitate data and specimen sharing and distributing knowledge amongst experts of various disciplines for the discovery of mechanisms and development of future interventions for prevention of type 1 diabetes. The ease in which data can be shared is critical to the success and mission of TEDDY. The reuse of data will also encourage further areas of study and support general knowledge sharing throughout the community. This data sharing plan is designed to enable the widest dissemination of data, while also protecting the privacy of the participants and the utility of the data, by de-identification and masking of potentially sensitive data elements. To support these objectives we will catalog and provide basic results sets through the NIDDK repository www.niddkrepository.org as well as create and deliver more comprehensive de-identified data sets, for each completed aspect of the TEDDY study. All applications for access of these data sets must be submitted through an application to NIDDK central repository as outlined on their web site www.niddkrepository.org. Access for the data submitted to the NIDDK Data Repository will be determined by the NIDDK. All investigators who receive TEDDY resources must agree to acknowledge the TEDDY Study and the NIDDK central repository. This approach is fully compliant with the NIH public data sharing policy (http://grants.nih.gov/grants/policy/data_sharing). That policy states that the NIH expects and supports the timely release and sharing of final research data from NIH-sponsored studies for use by other investigators and that the definition of “the timely release and sharing” to be no later than the acceptance for publication of the main findings from the final data set.

This policy is construed to include tissues, genetic material, and biological specimens as well as de-identified data from the TEDDY Data Coordinating Center (DCC). It is noted that biological specimens are now deposited with the NIDDK Repository, whereas, data is stored at the DCC.

The main findings of the TEDDY study are not anticipated before ten to fifteen years of follow-up of enrolled subjects plus the additional time it takes to complete assays of collected specimens. Therefore, it is the TEDDY policy that the members of the TEDDY Steering Committee shall retain control and primary rights to the data and biologic specimens developed and obtained under NIH funding, subject to Government rights of access consistent with current HHS, PHS, and NIH policies. This policy further establishes that this right shall cover all data and biological specimens deposited at the NIDDK repository or retained at any individual laboratory that received them under the auspices of TEDDY that contribute to achieving study outcomes. Such primary right shall exist until the conditions in paragraph 1, above, are fulfilled, namely no later than the acceptance for publication of the main findings from the final data set.

As the nature of the TEDDY study includes a nested case-control design in which early TEDDY samples may be selected for future assays, all samples and data will remain under the control of the TEDDY study group and will be made available to the scientific community following the established TEDDY policies governing ancillary studies (<http://teddy.epi.usf.edu/research/>).

TEDDY will be conducting analyses from the samples in two phases. The first phase will involve analysis of specific demographic, genetic, genomic, epigenetic, psychosocial, dietary, infectious, environmental, metabolomic, and proteomic measures when approximately 400 subjects develop

islet autoimmunity. The second phase of analyses will relate these measures to the development of islet autoimmunity and type 1 diabetes in, respectively, 800 and 400 subjects, at the conclusion of follow-up. In the meantime subjects will continue to be followed for islet autoimmunity and development of diabetes and further analysis of various measures and testing of other hypothesis will be conducted.

Thus data sets will be made available to the scientific community at the different time points when they do not affect the ability of the TEDDY group to achieve its aims. In making such data available through the NIDDK data repository, the TEDDY group will ensure that the data is cleaned, locked, and current as of a certain date. To be consistent with this idea, the following data sharing plan is to be put in place

Data description	Time-line for release to the NIDDK repository
Data used for a publication	3-6 months after the manuscript has appeared in print to allow time for publication of related manuscripts using the same dataset.
Screening data on all 424,788 subjects to include: HLA screening genotype, family history of diabetes, country of residence	6 months after the manuscript on TEDDY screening has appeared in print.
Baseline data on the 8677 subjects enrolled in TEDDY follow-up: demographics, HLA confirmation genotype, family history of diabetes	One year after the cleaned dataset becomes available.
All lab data such as SNP, microbiome and viral metagenomics, dietary biomarkers, metabolomics and gene expression	One year after the cleaned dataset becomes available.
Pilot studies data	One year after the cleaned dataset becomes available.

In addition, as the TEDDY study is committed to sharing data with the greater scientific community while the study is ongoing and to engage researchers in data analysis, the TEDDY Study has also developed a policy for the release of TEDDY data not in the NIDDK Repository. Please see this policy on the TEDDY public web site (<http://teddy.epi.usf.edu/research/>).