

Supporting Statement B for

Rapid Throughput Standardized Evaluation of  
Transmissible Risk for Substance Use Disorder in Youth  
(NIDA)

August 1, 2014

Name: Harold Perl, PhD  
Chief, Prevention Research Branch,  
Division of Epidemiology, Services & Prevention Research,  
National Institute on Drug Abuse,  
6001 Executive Blvd., Room 5163, MSC 9589,  
Bethesda, MD 20892-9589  
Telephone: 301-443-1942  
Fax: 301-443-2636  
Email: [hperl@nida.nih.gov](mailto:hperl@nida.nih.gov)

Table of Contents

<b>B.</b>	<b>COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS.....</b>	<b>3</b>
B.1	RESPONDENT UNIVERSE AND SAMPLING METHOD.....	3
B.2	PROCEDURES FOR THE COLLECTION OF INFORMATION.....	4
B.3	METHODS TO MAXIMIZE RESPONSE RATES AND DEAL WITH NON-RESPONSE.....	9
B.4	TEST OF PROCEDURES OR METHODS TO BE UNDERTAKEN.....	10
B.5	INDIVIDUALS CONSULTED ON STATISTICAL ASPECTS AND INDIVIDUALS COLLECTING AND/OR ANALYZING DATA.....	11

## **B.1 Respondent Universe and Sampling Method**

### Part 1: Standardization of the TLI (School Survey)

The potential universe of schools includes all public and private high schools in the continental United States; as of 2008, this approximate N equaled 132,446 (source: National Center for Education Statistics, Digest of Education Statistics, Table 5. Number of educational institutions, by level and control of institution: Selected years, 1980-81 through 2008-09).

The potential respondent universe includes all 14 to 18 year olds in the continental United States enrolled in a public or private high school (approximate N=24,412,000; source: U.S. Census Bureau, Current Population Study, Table 2. Single Grade of Enrollment and High School Graduation Status for People 3 Years Old and Over, by Sex, Age (Single Years for 3 to 24 Years), Race, and Hispanic Origin; October 2011)

Response Rate: The school level response rate will fall between 30% to 40% of the schools initially randomly selected from the narrowed list. The study will use a sampling with replacement strategy. Replacement schools will be selected based on similar racial/ethnic school population composition as those from the initial round of selection that declined participation. In this way, replacement schools will be substantially similarly as population-representative as the initially selected school. Rutgers/BCSR achieved a 73% rate in New Jersey for the 2013 YRBS. Among the 42 states that obtained weighted data for that survey administration, response rates ranged from 60% to 94%.

([http://www.cdc.gov/mmwr/preview/mmwrhtml/ss6304a1.htm?s\\_cid=ss6304a1\\_e](http://www.cdc.gov/mmwr/preview/mmwrhtml/ss6304a1.htm?s_cid=ss6304a1_e))

Thus, based on BCSR's extensive prior experience conducting the CDC-sponsored Youth Risk Behavior Survey for the New Jersey Department of Education, the student-level response rate is predicted to fall between 70% to 75%.

## Part 2: Psychometric Validation of the TLI (Twins Survey)

The potential universe includes all sets of identical and fraternal twins in the U.S. Per the CDC/NCHS, National Vital Statistics System, the rate of twin births between 1995 and 1999 grew from 20 to 25 births per 1000. Between 1995 and 1999 there were approximate 20 million live births (source: Department of Health and Human Services, National Center for Health Statistics). This equates to approximately 400,000 to 500,000 sets of twins ages 14-18 years old in the United States.

Response Rate: The twins survey is conducted on a convenience sample, rather than a probability sample. As such, there is no calculable response rate. However, under these circumstances, convenience sampling is substantially more facile and less expensive than probability sampling. Because the nature of the research question under investigation, i.e., the existence of a genetic component of addiction propensity, is a function of the biological traits of the sampling unit, rather than a randomly distributed population characteristic, there is no justifiable purpose in generating sample estimates of population parameters. Rather, the focus is on determining whether the phenomenon of inheritable addiction propensity occurs across two categories of biological relationships, i.e., identical twins and fraternal twins. In the absence of the need for population estimated, the cost to the federal

government of probability sampling (and, in turn, a calculable response rate) cannot be justified.

## **B.2 Procedures for the Collection of Information**

### Part 1: Standardization of the TLI (School Survey)

A sample of 36 high schools across the continental United States will be used. A list of all public and private high schools within each of the 9 Census Division will be obtained from a commercial sampling company.

The first stratum categorizes the sample of all high schools in the continental United States by Census Division. The primary concern is to ensure relevant variation to capture regional differences in attitudes and behaviors relating to the predictive cofactors for substance abuse disorder. Population size differences will be accounted for by post-stratification weighting.

The second stratum involves selecting schools with race/ethnicity distributions statistically similar to the race/ethnicity population distributions for those Census Divisions; from that “similar distribution” list, a set of schools will be randomly selected. Using the most recently available Census Bureau data, the racial/ethnic population distributions will be determined for each Division. For the school sample frame in each Division, the racial/ethnic distribution will be determined either by using Census figures for the primary municipality in the school distribute or, optimally, by using the school’s own race/ethnicity distribution data if it agrees with the Census racial and ethnic categories. From the master,

commercially-obtained sample frame, the principal investigators will select schools that best mirror the racial/ethnic distribution for the Division. Schools on that narrowed list will be rank-ordered by student enrollment for grade 8-12 or 9-12, and a random selection method, with the odds of selection set to be proportional to the size of the school, will be employed. In this way, if, for example, a school represents .02% of the student population in Division 1, the probability of selection for that school will be .02%. Sampling will be “with replacement” such that if, after the initial recruitment effort fails to produce 4 participating schools, a second iteration of the process will be conducted using the full sample. Table 4, below, shows the school sampling by population distribution.

The third stratum involves sampling students within schools. As noted above, a list of all classes (and number of students per class) that includes students in the appropriate grades will be obtained from the school liaison. From that list, a sampling ratio will be calculated so that the total number of students sampled from the school will meet the overall targeted sample for that school. The appropriate number of classes will then be randomly selected at between 4 to 10 classes from any school. The completed survey goal for the school is 130 to 160 completes. As such, the range of total completed surveys per region will be 520 to 640; and the range of total completed surveys for the sample will be between 4,680 to 5,760, which is adequate to populate each of the 30 cells, with 50 or more students, as shown in Table 5, below.

The construction of sampling cells in Table 5 is the product of stratification across age cohorts (14, 15, 16, 17 and 18 year olds), gender (male and female), the five OMB race

categories (White, Black or African-American, Asian , Native Hawaiian-Other Pacific Islander, American Indian-Alaska Native), with the last three combined for sampling pragmatics, and ethnicity (Hispanic or Latino and not Hispanic or Latino). Overall, 40 cohorts will be generated across four demographic categories of race and ethnicity with a “completed interview goal” of a minimum of  $n=50$  for each cell White, Black/African-American, and Hispanic and 25 for Asian/Other. An example of cell fulfillment would, for example, be 50 completed interviews for the group “15 year olds-female-African-American.”

To properly demonstrate reliability and validity, a minimum of 1,500 completes must be collected for the study. The range of students, i.e., from 520-640 is, under the statistical laws of probability sampling, adequate to obtain a representative sample of the population distribution within the nine regional classification. In other words, to illustrate, a sample of 600 students in the smallest region (New England) will return a margin of sampling error at 95% at the 45% to 55% marginal range of  $\pm 3.98\%$ . Due to the stable mathematical principles of sampling theory, the same size sample, i.e., 600 students, under the same parameters, will also return a margin of sampling error of  $\pm 3.98\%$  in the largest region (South Atlantic). Table 2 provides an overview of the sample quotas.

The degree of accuracy in sampling required is an empirical question, which is a function of the degree of similarity between the race/ethnicity distributions of the schools in the statistical sample and the race/ethnicity distributions of the populations for the Census Divisions in which those schools are situated.

Part 2: Psychometric Validation of the TLI (Twins Survey)

A total of three surveys (the Demographic and Drug and Alcohol survey, the Dysregulation Inventory survey, and the TLI survey) will be administered to 150 sets of identical and 150 sets of fraternal twins in each of two consecutive years. The total completed sample size is 600 youths represented, as much as is possible, equally by each gender and evenly distributed within the 14-18 age range. To find the sample of 14-18 year old twins online, parents (of twins) will be recruited via the Pittsburgh Registry of Infant Multiplets (PRIM), the National Organization of Mothers of Twins Clubs, Inc., and/or a commercial online twins registry.

**Table 4: School Sampling by Population Distribution**

<b>Region</b>	<b>Division</b>	<b>States</b>	<b>2012 Census Population (14 to 18 year olds)</b>	<b>% of 2012 Population</b>	<b>Target Schools</b>	<b># of students</b>
1 Northeast	1 New England	Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut	966,624	4.59%	4	520-640
	2 Mid-Atlantic	New York, Pennsylvania, New Jersey	2,699,296	12.81%	4	520-640
2 Midwest	3 East North Central	Wisconsin, Michigan, Illinois, Indiana, Ohio	3,186,571	15.12%	4	520-640
	4 West North Central	Missouri, North Dakota, South Dakota, Nebraska, Kansas, Minnesota, Iowa	1,379,793	6.55%	4	520-640
3 South	5 South Atlantic	Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida	3,920,480	18.60%	4	520-640
	6 East South Central	Kentucky, Tennessee, Mississippi, Alabama	1,233,884	5.86%	4	520-640



4 West	7 West South Central	Oklahoma, Texas, Arkansas, Louisiana	2,638,176	12.52%	4	520-640
	8 Mountain	Idaho, Montana, Wyoming, Nevada, Utah, Colorado, Arizona, New Mexico	1,543,869	7.33%	4	520-640
	9 Pacific	Washington, Oregon, California, Hawaii, Alaska	3,504,588	16.63%	4	520-640
TOTAL			21,073,281	100.00%	36	4680-5760

**Table 5: Minimum Sample Quotas (cells contain intended number of completed interviews)**

		Gender and Race/Ethnicity							
		Male				Female			
		White	Black	Hispanic	Asian/Other	White	Black	Hispanic	Asian/Other
AGE	14	50	50	50	25	50	50	50	25
	15	50	50	50	25	50	50	50	25
	16	50	50	50	25	50	50	50	25
	17	50	50	50	25	50	50	50	25
	18	50	50	50	25	50	50	50	25

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

#### Part 1: Standardization of the TLI (School Survey)

To maximize response rates, the data collection team uses protocols developed and tested in the Youth Risk Behavior Survey/Student Health Survey for the New Jersey Department of Education and the Risk and Protective Factor Survey for the New Jersey Department of Health. As is our standard school recruitment protocol, school will be offered an institutional-level incentive as part of the response rate maximization protocol. Several principles inform response rate maximization:

- (1) engaging personal contact with superintendents and principals;

- (2) mailing of the initial contact packet to schools using Postal Service Priority mail;
- (3) engaging in timely follow-up calls after the initial study mailing;
- (4) conveying a clear understanding of IRB and OMB approval of the data collection effort;
- (5) conveying a clear understanding of the law and regulations regarding parental consent under the relevant state law;
- (6) conveying a clear understanding of the voluntary nature of the data collection effort; and,
- (7) providing a school-level incentive.

The initial mailing packet includes a recruitment letter from the study sponsor that denotes the importance of the study and its goals; also included are a study fact sheet, a school agreement form, a sample student/parental consent form, and a printed version of the online questionnaire.

#### Part 2: Psychometric Validation of the TLI (Twins Survey)

To maximize response rates of the twins sample, the following techniques will be employed:

- Simple administration: Instructions will be clear and focused on online survey completion.
- Upfront privacy notice: Respondents will be informed of confidentiality (to the extent permitted by law) of results and that data will only be used in aggregate.
- Reminder of survey: Respondents will receive up to 4 notifications to complete the survey.
- Reasonable incentive: Respondents will receive a fair and commensurate incentive (\$10) for their participation in completion of the online survey.

#### **B.4 Test of Procedures or Methods to be Undertaken**

The TLI survey instruments has been tested in small focus group (N<9) for reading comprehension, technology use, female use and African American use. Upon approval it will also be thoroughly tested by BCSR, EXACT, and CEDAR staff. The staff selected for pretesting will be “strangers” to the project so as to best minimize any familiarity effect and better ensure independence of assessment and feedback. In specific, the instrument will be tested to assure accessibility, functionality, accuracy of data recording, and ease of use.

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

Center for Education and Drug Abuse Research (CEDAR); University of Pittsburgh:

Levent Kirisci            412-864-2461  
Maureen Reynolds      412-488-5006  
Michael Vanyukov       412-864-2458

EXACT Sport LLC (EXACT):

Barry Tarter             312-854-2352  
Simon Clements         312-854-2356

Bloustein Center for Survey Research (BCSR); Edward J. Bloustein School of Planning and Public Policy, Rutgers, The State University of New Jersey:

Christopher Bruzios     848-932-2778  
Marc Weiner             848-932-2765  
Timothy Mackinnon     848-932-2775  
Siobhan Foley           848-932-2782  
Orin Puniello            848-932-2767