APPENDIX A

SAMPLING, DATA COLLECTION, AND ANALYSIS PROCEDURES: NLTS WAVE 1 PARENT INTERVIEW/SURVEY AND NLTS2 WAVE 2 PARENT-YOUTH INTERVIEW/SURVEY

APPENDIX A
SAMPLING, DATA COLLECTION, AND ANALYSIS PROCEDURES: NLTS WAVE 1 PARENT INTERVIEW/SURVEY AND NLTS2 WAVE 2 PARENT-YOUTH INTERVIEW/SURVEY

This appendix describes several aspects of the NLTS and NLTS2 methods relevant to the data reported in this document and to comparisons between the studies, including:

- Sampling of local education agencies (LEAs), schools, and students
- Parent interview and survey procedures and response rates
- Weighting of the parent interview/survey data
- Analytic adjustments to increase the comparability of the study samples
- Estimation and use of standard errors
- Unweighted and weighted sample sizes
- Calculation of statistical significance
- Measurement issues.

Overview of the NLTS and NLTS2 Samples

The samples for both studies were constructed in two stages. A stratified random sample of LEAs was selected from the universe of operating LEAs that served students receiving special education in at least one grade from 7th through 12th grades in the 1983-84 and 1999-2000 school years. These LEAs and all state-supported special schools that served primarily students with hearing and vision impairments and multiple disabilities were invited to participate in the study. Targets of recruiting 400 and 497 participating LEAs were set for the two studies, respectively, and as many special schools as possible. From these would be selected target student samples of about 14,000 (NLTS) and 12,000 students (NLTS2). Approximately three-fourths of the target number of LEAs was reached in NLTS and 101% in NLTS2.

For both studies, the roster of all students receiving special education from each participating LEA[1] and special school was stratified by disability category (11 in use in 1987 and 12 in 2000) and age. Students then were selected randomly from each disability category and age group. Sampling fractions were calculated that would produce enough students in each category so that, in the final year of each study, findings would generalize to most categories individually with an acceptable level of precision, accounting for attrition and for response rates to the parent/youth interview. A total of 10,369 and 11,276 students were selected and eligible to participate in the NLTS and NLTS2 parent interview/surveys, respectively.

Details of the LEA and student samples are provided below.

---

[1] LEAs were instructed to include on the roster any student for which they were administratively responsible, even if the student was not educated within the LEA (e.g., attended school sponsored by an education cooperative or was sent by the LEA to a private school). Despite these instructions, some LEAs may have underreported students served outside the LEA.

The LEA Samples

Defining the Universe of LEAs

The NLTS and NLTS2 samples include only LEAs that had teachers, students, administrators, and operating schools—that is, "operating LEAs." They exclude such units as supervisory unions; Bureau of Indian Affairs schools; public and private agencies, such as correctional facilities; LEAs from U.S. territories; and LEAs with 10 or fewer students in the NLTS2 age range, which would be unlikely to have students with disabilities.

The public school universe data file maintained by Quality Education Data (QED) for 1998 was used to construct the NLTS2 sampling frame because it had more recent information than the alternative list maintained by the National Center for Education Statistics (NCES). For NLTS, a combination of QED and NCES data was used for the 1983 and 1984 school years, respectively. In NLTS, a sample of 1,600 LEAs was surveyed by telephone to collect data on LEAs for sample and bias estimation purposes. (Details of the NLTS Wave 1 sample can be found in Javitz & Wagner, 1990.) Correcting for errors and duplications resulted in a master list of 13,180 (NLTS) and 12,435 (NLTS2) LEAs that met the selection criteria for the two studies. These comprised the LEA sampling frames.

Stratification

The LEA samples were stratified to increase the precision of estimates, to ensure that low-frequency types of LEAs (e.g., large urban districts) were adequately represented in the samples, to improve comparisons with the findings of other research, and to make the studies responsive to concerns voiced in policy debate (e.g., differential effects of federal policies in particular regions, LEAs of different sizes). Three stratifying variables were used:

Region. This variable captures essential political differences, as well as subtle differences in the organization of schools, the economic conditions under which they operate, and the character of public concerns. The regional classification that was used by the U.S. Department of Commerce, the U.S. Bureau of Economic Analysis, and the National Assessment of Educational Progress was selected (categories are Northeast, Southeast, Midwest, and West).

LEA size (student enrollment). LEAs vary considerably by size, the most useful available measure of which is student enrollment. A host of organizational and contextual variables are associated with size, and they exert considerable potential influence over the operations and effects of special education and related programs. In addition, total enrollment serves as an initial proxy for the number of students receiving special education in an LEA. The QED database provides enrollment data from which LEAs were sorted into the following categories:[2]

NLTS

- **Huge** (enrollment of 50,000 or more).
- **Very large** (enrollment of 25,000 to 49,999).
- **Large (**enrollment of 10,000 to 24,999).

[2] NLTS size strata were determined by logical dividing points using multiples of 500 students. NLTS2 strata are quartiles.

- **Medium (**enrollment of 2,500 to 9,999).
- **Small** (enrollment of 500 to 2,499).
- **Very small** (enrollment less than 500).

NLTS2

- **Very large** (estimated[3] enrollment greater than 14,931 in grades 7 through 12).
- **Large** (estimated enrollment from 4,661 to 14,931 in grades 7 through 12).
- **Medium** (estimated enrollment from 1,568 to 4,660 in grades 7 through 12).
- **Small** (estimated enrollment from 11 to 1,567 in grades 7 through 12).

LEA/community wealth. As a measure of district wealth, the Orshansky index (the proportion of the student population living below the federal definition of poverty, Employment Policies Institute, 2002) is a well-accepted measure. The distribution of Orshansky index scores was organized into four categories of LEA/community wealth, as follows:[4]

NLTS

- **High** (0 to 4% disadvantaged youth).
- **Medium** (5% to 9% disadvantaged youth).
- **Low** (10% to 19% disadvantaged youth).
- **Very low** (20% or more disadvantaged youth).

NLTS2

- **High** (0% to 13% disadvantaged youth).
- **Medium** (14% to 24% disadvantaged youth).
- **Low** (25% to 43% disadvantaged youth).
- **Very low** (43% or more disadvantaged youth).

The three variables generated 96- and 64-cell grids for the two studies, into which the universes of LEAs were arrayed.

LEA Sample Size

On the basis of an analysis of LEAs' estimated enrollment across LEA size, and estimated sampling fractions for each disability category, targets of 400 and 497 LEAs (and as many state-sponsored special schools as would participate) were considered sufficient to generate the student samples needed for the two studies (Exhibit A-1). Taking into account expectations

---

3 Enrollment in grades 7 through 12 was estimated by dividing the total enrollment in all grade levels served by an LEA by the number of grade levels to estimate an enrollment per grade level. This value was then multiplied by 6 to estimate the enrollment in grades 7 through 12.

4 NLTS wealth strata were defined by logical divisions, with strata being multiples of 5 percentage points. NLTS2 strata are quartiles.

regarding the rate at which LEAs would refuse to participate (which experience in the intervening years suggests would be dramatically higher in 2000 than in 1987), samples of 628 and 3,635 LEAs were invited to participate in the two studies, respectively. A total of 303 and 501 LEAs provided students for the two study samples—76% and 101% of the target numbers needed and 48% and 14% of those invited. Analyses of the region, size, and wealth of the LEA sample, both weighted and unweighted, confirmed that that the weighted LEA sample closely resembled the LEA universe with respect to those variables. However, in addition to ensuring that the LEA sample matched the universe of LEAs on variables used in sampling, it was important to ascertain whether the stratified random sampling approach resulted in skewed distributions on relevant variables not included in the stratification scheme. Thus, additional extensive analyses were conducted on the LEA sample of both studies.

Exhibit A-1
FIRST STAGE SAMPLE SIZES

	NLTS	NLTS2
Target LEA sample sought	400	497
Sample invited to participate		
LEAs	628	3,635
Special schools	84	77
TOTAL	712	3,712
Sample participating		
LEAs	303	501
Special schools	22	
TOTAL	325	
Percentage of invited		
LEAs	48%	14%
Special schools	26%	
TOTAL	46%	
Percentage of LEA target	**76%**	**101%**

NLTS analyses involved comparing the 303 participating LEAs with a sample of 1,600 LEAs randomly selected from the universe of LEAs and contacted in a brief telephone survey. The only significant or meaningful difference found between the NLTS sample and the larger survey sample was that NLTS underrepresented students in LEAs that served grades kindergarten through eighth grade. It was hypothesized at the time that K-8 districts may not have perceived themselves to be secondary districts and refused to participate at higher rates because only their seventh and eighth grade students would have met the sample criteria. No variables, beyond those used to stratify the sample, were used in constructing weights at the LEA level.

NLTS2 analyses involved several stages. The first involved selecting three variables from the QED database on which to compare the "fit" between the first-stage sample and the population: the LEA's racial/ethnic distribution of students, the proportion who attended college, and the urban/rural status of the LEA. This analysis revealed that the sample of LEAs somewhat underrepresented African American students and college-bound students, and overrepresented Hispanic students and LEAs in rural areas. Thus, in addition to accounting for stratification variables, LEA weights were calculated to achieve a distribution on the urbanicity and racial/ethnic distributions of students who matched the universe.

To determine whether the resulting weights, when applied to the participating NLTS2 LEAs, accurately represented the universe of LEAs serving the specified grade levels, data collected from the universe of LEAs by the U.S. Department of Education's Office of Civil Rights (OCR) and additional items from QED were compared for the weighted NLTS2 LEA sample and the universe. Finally, the NLTS2 participating LEAs and a sample of 1,000 LEAs that represented the universe of LEAs were surveyed to assess a variety of policies and practices known to vary among LEAs and to be relevant to secondary-school-age youth with disabilities.

Analyses of both the extant databases and the LEA survey data confirm that the weighted NLTS2 LEA sample accurately represents the universe of LEAs.

The Student Samples

Determining the size of the NLTS and NLTS2 student samples took into account the duration of the study (5 and 10 years, respectively), desired levels of precision, and assumptions regarding attrition and response rates. (Obviously, these kinds of assumptions for NLTS were not informed by the experience gleaned from it and other longitudinal studies conducted in the intervening years.) The studies' sample designs called for findings to be generalizable to students receiving special education as a whole and for each of the special education disability categories in use at the time. Standard errors were to be no more than 3.2% and 3.6% for the two studies, respectively, except for the low-incidence categories. Assuming a 50% sampling efficiency, analyses for the two studies determined that approximately 13,000 and 12,000 students would need to be sampled to ensure sufficient youth would have a parent/youth interview in the final wave of each study.

LEAs and special schools were contacted to obtain their agreement to participate in the study and to request rosters of students receiving special education. NLTS sampled students ages 13 to 21, and NLTS2 sampled students ages 13 through 16. For both studies, students had to have been in at least 7th grade.[5] Requests for rosters for both studies specified that they contain the names of students receiving special education under the jurisdiction of the LEA, the disability category of each student, and the students' birth dates or ages. NLTS also requested the name of students' schools. NLTS2 requested that student addresses and telephone numbers be included on rosters; this information was obtained in a second contact with LEAs for NLTS. Some LEAs in both studies would provide only identification numbers for students, along with the corresponding birth dates and disability categories. When students were sampled in these LEAs, identification numbers of selected students were provided to the LEA, along with materials to mail to their parents/guardians (without revealing their identity).

After estimating the number of students receiving special education in the NLTS2 age range, the appropriate fraction of students in each category was selected randomly from each LEA and special school. In cases in which a family had more than one child included on a roster, only one was eligible to be selected. LEAs and special schools were notified of the students selected, and contact information for their parents/guardians was requested if it had not been provided initially.

Interviews/Surveys of Parents and Youth

The data source for the NLTS findings reported here was parents/guardians of NLTS sample members, who were interviewed by telephone or surveyed by mail in 1987.[6] NLTS2 data come from 2003 interviews with parents/guardians of NLTS2 sample members and of youth themselves when they were able to respond; youth who could not respond by telephone but could complete a self-administered questionnaire were mailed one.

---

[5] Students who were designated as being in ungraded programs also were sampled if they met the age criteria.

[6] More details of the NLTS data collection procedures are found in Wagner, Newman, & Shaver (1989).

Letters were sent to parents of youth in each study sample to notify them that their child had been selected for study participation and that an interviewer would attempt to contact them by telephone. The letters for both studies included a toll-free telephone number for parents to call to be interviewed if they did not have a telephone number where they could be reached reliably or if they wanted to make an appointment for the interview at a specific time.

Computer-assisted telephone interviewing (CATI) was used for both NLTS and NLTS2. NLTS interviews were conducted between June and September 1987, and parents who could not be reached by telephone were mailed a self-administered questionnaire. The questionnaire contained a subset of key items from the telephone interview. The questionnaire also requested a telephone number, and if a new working number was provided, a telephone interview also was attempted in an extended interview period through November 1987. NLTS2 interviews were conducted between May and December 2003. Only youth whose households included an adult member who spoke English or Spanish were included in NLTS and NLTS2 interviews. In the two studies, 96% and 97% of the interviews, respectively, were conducted in English. Out of 10,369 eligible youth in NLTS, 6,852 respondents provided data. Out of 8,210 eligible youth in NLTS2, data was provided for 6,888 youth.

Wave 2 Parent/Youth Interviews

NLTS2 sample members for whom working telephone numbers and addresses were available were eligible for the Wave 2 parent/youth telephone interview in 2003. Database matching procedures were used to maximize the eligible sample, as in Wave 1. Contact procedures alerting parents of the interviews also were similar for the two waves. The major distinction between the data collection methods in Waves 1 and 2 is that interviews were sought both with parents of NLTS2 sample members and with the youth themselves if they were able to respond to questions.

The first interview contact was made with parents of eligible sample members. Those who agreed to participate were interviewed using CATI. Items in this portion of the interview, referred to as Parent Part 1, focused on topics for which the parent was considered the most appropriate respondent (e.g., services received, family expectations, and support). At the end of Parent Part 1, the respondent was asked the following:

> *My next questions are about jobs (YOUTH'S NAME) may have had, schools (he/she) may have gone to, and about (his/her) feelings about (him/herself) and (his/her) life. The questions are similar to those I've been asking you, where (he/she) will be asked to answer using scales, like "very well," "pretty well," "not very well," or "not at all well." The interview would probably last about 20 to 30 minutes. Do you think that (YOUTH'S NAME) would be able to accurately answer these kinds of questions over the telephone?*

If youth could answer questions by phone, they also were told:

> *I also have some questions about (his/her) involvement in risk behaviors, like smoking, drinking, and sexual activity. Is it all right for me to ask (YOUTH"S NAME) questions like that?*

If parents consented, interviewers asked to speak with the youth or asked for contact information to reach the youth in order to complete the youth portion of the interview, referred to as Youth Part 2.

Parents who reported that youth could not answer questions by telephone were asked:

Would (he/she) be able to accurately answer these kinds of questions using a written questionnaire?

If parents indicated youth could complete a written questionnaire, they were asked for the best address to which to send a questionnaire, and a questionnaire was sent. The questionnaire contained a subset of items from the telephone interview that were considered most important for understanding the experiences and perspectives of youth. Multiple follow-up phone or mail contacts were made to maximize the response rate for the mail survey. Data from the mail survey and Youth Part 2 of the telephone interview were merged for analysis purposes.

If parents reported that youth could not answer questions either by telephone or written questionnaire or declined to have youth asked questions related to risk behaviors, interviewers asked them to continue the interview, referred to as Parent Part 2. If youth were reported to be able to complete a telephone interview or a written questionnaire but did not after repeated attempts, parents were contacted again and asked to complete Parent Part 2 in lieu of Youth Part 2.

Exhibit A-2 reports the sample members for whom there are data from the Wave 2 Parent Part 1 and Parent Part 2 telephone interviews and the Youth Part 2 telephone/mail survey.

Combining Parent and Youth Data

Youth Part 2 of the interview contains many items that were asked only of youth because they focus on youth's perspectives or attitudes (e.g., job satisfaction, self-concept). However, the majority of the interview items that were in Youth Part 2 also were included in Parent Part 2 so that data would be available for them, regardless of whether a parent or youth completed the interview or the mail questionnaire. Thus, in preparing the data for analysis, responses to these overlapping items from parents and youth were combined—i.e., data for many Part 2 items combine responses from parents and youth in the proportions with which they completed Part 2 of the interviews, indicated in Exhibit A-2.

Exhibit A-2
RESPONSE RATES FOR NLTS2 WAVE 2 PARENT/YOUTH DATA COLLECTION

	Number	Percentage
Total eligible sample	8,210	100.0
Respondents		
Completed Parent Part 1 telephone interview	6,888	83.9
Completed Parent Part 2 telephone interview	2,997	36.5
Completed Youth Part 2 telephone interview or mail questionnaire	3,375	41.1
Completed Part 1 and either Parent or Youth Part 2	6,372	77.6
Total nonrespondents (no parent or youth data)	1,322	16.1

There also is a relatively small set of items that appeared in Parent Part 1 as well as Youth Part 2. These were considered critical variables for which the maximum amount of data would be needed; they were included in Part 1 to avoid the risk that a Part 2 would not be completed with either the parent or youth. However, a small number of these also were included in Youth Part 2 because a youth was potentially the more knowledgeable respondent. In such cases, the youth response was used when combining parent and youth data.

Combining data across respondents raises the question of whether parent and youth responses would concur—i.e., would the same findings result if parent responses were reported instead of youth responses. Exhibit A-3 reports the level of congruence in parent and youth responses to four items related to key outcomes of interest. However, a high degree of congruence gives confidence that accurate information is being collected, regardless of who provided Part 2 responses.

When both parents and youth were asked whether the youth belongs to an organized community group, currently works for pay, and worked for pay in the past 2 years, and whether currently employed youth earned less than $5.15 per hour, $5.15 to $6.00 per hour, $6.01 to $7.00 per hour, or more than $7.00 per hour, their responses agreed from 68% to 87% of the time. The greatest congruence (87%) is noted regarding youth's current employment, with high congruence (79%) also evident regarding wages for that employment. There is somewhat less, although still relatively high, agreement regarding employment in the preceding 2 years (74%). Parents and youth were least likely to agree on whether youth belonged to an organized group in the community. This item could be expected to have greater discrepancy than those dealing with employment because parents could be less aware of youth's social or leisure time activities than of employment, the evidence of which would be visible in the wages earned and spent.

Exhibit A-3
CONGRUENCE OF PARENT AND YOUTH RESPONSES TO KEY ITEMS

	Percentage with:		
	Congruent Responses	Parent Answering Yes/ Higher Wage, Youth No/ Lower Wage	Parent Answering No/Lower Wage, Youth Yes/Higher Wage
Youth currently working for pay	86.9	5.7	7.5
Current hourly wage	79.1	5.5	15.4
Youth worked for pay in past 2 years	73.6	7.6	18.9
Youth belongs to an organized group in the community	68.5	4.4	27.1

It is impossible to determine the cause of discrepant responses. Complete congruence would not be expected, even with both respondents answering accurately, because Parent Part 1 could have been completed well before the subsequent Part 2 interview during the 7-month interview period; the status of youth could have changed in the intervening period. In such cases, both responses would be accurate at the time given. However, discrepancies also could result from one response being inaccurate, either because a respondent gave a socially desirable response (e.g., reported a youth was employed when he or she was not) or because the respondent (usually the parent) had inaccurate information (e.g., a youth no longer living with a parent had not informed the parent regarding a community group he or she had joined, leading to a negative parent response regarding group membership when a positive response was accurate). Although it is not possible to tell which of two discrepant responses is correct, it is noteworthy that with the exception of current employment, discrepant cases are more likely to result from a positive response from youth when parents responded negatively (e.g., youth reported higher wages or a higher rate of group membership than parents). Thus, for some items, youth for whom data were collected through Youth Part 2 may appear to have more positive experiences than those for whom data were collected through Parent Part 2 because of the source of the data, in addition to or instead of actual differences in their experiences. Again, this difference does not necessarily imply inaccuracies in the data, but it does affirm the difference in the knowledge and perspectives of parents and youth.

Weighting the Wave 1 Parent Data

The percentages and means reported in the data tables are estimates of the true values for the population of youth who had been out of school up to 2 years. The estimates are calculated from responses of parents of NLTS and NLTS2 sample members. The response for each sample member is weighted to represent the number of youth in his or her disability category in the kind of LEA (i.e., region, size, and wealth) or special school from which he or she was selected.

Exhibit A-4 illustrates the concept of sample weighting and its effect on percentages or means that are calculated for students with disabilities as a group. In this example, 10 students are included in a sample, 1 from each of 10 disability groups, and each has a hypothetical value regarding whether that student participated in organized group activities outside of school (1 for yes, 0 for no). Six students participated in such activities, which would result in an unweighted value of 60% participating. However, this would not accurately represent the national population of students with disabilities because many more students are classified as having a learning disability than orthopedic or other health impairments, for example. Therefore, in calculating a population estimate, weights in the example are applied that correspond to the proportion of students in the population that are from each disability category. (Actual study weights account for several aspects of the students and the districts from which they were chosen.) The sample weights for this example appear in column C. Using these weights, the weighted population estimate is 87%. The percentages in NLTS and NLTS2 are similarly weighted population estimates, whereas the sample sizes are the actual number of cases on which the weighted estimates are based (similar to the 10 cases in Exhibit A-4).

Exhibit A-4
EXAMPLE OF WEIGHTED PERCENTAGE CALCULATION

Disability Category	A Number in Sample	B Participated in Group Activities	C Example Weight for Category	D Weighted Value for Category
Learning disability	1	1	5.5	5.5
Speech/language impairment	1	1	2.2	2.2
Mental retardation	1	1	1.1	1.1
Emotional disturbance	1	0	.9	0
Hearing impairment	1	1	.2	.2
Visual impairment	1	1	.1	.1
Orthopedic impairment	1	0	.1	0
Other health impairment	1	1	.6	.6
Autism	1	0	.2	0
Multiple disabilities	1	0	.1	0
TOTAL	10	6	10	8.7

Unweighted sample percentage = 60% (Column B total, divided by Column A total)

Weighted population estimate = 87% (Column D total, divided by Column C total)

The students in LEAs and state schools with parent interview/survey data were weighted to represent the universe of students in LEAs and state schools at the two study time points. NLTS weighting procedures are detailed in Javitz & Wagner (1990). NLTS2 used the following process:

- For each of the 64 LEA sampling cells, an LEA student sampling weight was computed. This weight is the ratio of the number of students in participating LEAs in that cell, divided by the number of students in all LEAs in that cell in the universe of LEAs. The weight represents the number of students in the universe who are represented by each student in the participating LEAs. For example, if participating LEAs in a particular cell served 4,000 students and if the universe of LEAs in the cell served 400,000 students, the LEA student sampling weight would be 100.
- For each of the 64 LEA cells, the number of students in each disability category was estimated by multiplying the number of students with that disability on the rosters of participating LEAs in a cell by the adjusted LEA student sampling weight for that cell. For example, if 350 students with learning disabilities were served by LEAs in a cell, and the LEA student sampling weight for that cell was 100 (i.e., each student in the sample of participating LEAs in that cell represented 100 students in the universe), estimates would suggest 35,000 students with learning disabilities in that cell in the universe.
- For the state schools, the number of students in each disability category was estimated by multiplying the number of students with that disability on the rosters by the inverse of the proportion of state schools that submitted rosters.

- The initial student sampling weights were adjusted by disability category so that the sum of the weights (i.e., the initial student sampling weights, multiplied by the number of students for whom interviews were completed) was equal to the number of students in the geographical and wealth cells of each size strata. The adjustments were typically small and essentially served as a nonresponse adjustment. However, the adjustments could become substantial when there were relatively few interviewees (as occurred in the small and medium strata for the lowest incidence disabilities) because in these cases, some cells might not have any interviewees, and it was necessary to adjust the weights of other interviewees to compensate. Two constraints were imposed on the adjustments: (1) within each size stratum, the cell's weights could not vary from the average weight by more than a factor of 2, and (2) the average weight within each size stratum could not be larger than 4 times the overall average weight. These constraints substantially increased the efficiency of the sample at the cost of introducing a small amount of weighting bias (discussed below).
- In a final step, the weights were adjusted so that they summed to the number of students in each disability category, as reported to OSEP by the states for the 2000-2001 school year (OSEP, 2001).

The imposition of constraints on the adjusted weights increased sampling efficiency at the cost of introducing a small amount of bias. The average efficiency increased from 51.7% to 67.4%; the largest increases in sampling efficiency occurred for youth with emotional disturbances (from 44.4% to 81.0%) and for those with multiple disabilities (from 32.1% to 56.8%). Biases introduced by the imposition of constraints on the student weights generally were very small. The largest bias in size distribution was for youth with visual impairments (decreasing from 17.1% in the smallest size stratum to 11.6%) and those with autism (decreasing from 21.3% in the smallest size stratum to 17.5%). All other changes in the size distribution were 1.5% or less, and the average absolute change was only 0.4%. The largest bias in wealth distribution was for those with multiple disabilities (from 22.2% in wealth stratum 3 to 16.6%, and from 18.3% in wealth stratum 4 to 22.0%). All other changes were 2.1% or less, and the average absolute change was only 0.6%. All biases in regional distribution were 2.1% or less, and the average absolute change was only 0.5%. Considering the increase in sampling efficiency, these biases are considered acceptable.

The reason for the reduction in the proportion of students represented in the cells mentioned above is that there were relatively few students with interview/survey data in those cells. For example, small LEAs had only 21 students with visual impairments with data, requiring that they represent an estimated 1,701 students with visual impairments from small LEAs. The weighting program determined that the average weight required (i.e., 81.0) violated the constraints, and therefore reduced these weights to a more reasonable value (i.e., 56.2).

Analytic Adjustment to Increase the Comparability of Study Samples

The NLTS and NLTS2 samples are similar in many respects. Yet, they differ in important ways that make a comparison between youth in the full samples of the two studies inadvisable because misleading conclusions could be drawn from such comparisons. One important distinction is the age of youth in the two studies. NLTS includes youth who were ages 13 to 21 when selected and 15 to 23 when the Wave 1 parent data were collected. NLTS2, in contrast, includes youth who were 13 to 16 when selected and 15 to 19 when Wave 2 parent data were collected. Thus, the full sample of youth with NLTS Wave 1 parent data included youth who were older than any in NLTS2 (20- through 23-year-olds), and NLTS2 included youth who were younger than any included in NLTS (13- and 14-year-olds). Because age is such a powerful determinant of the experience of adolescents, comparisons made in this report between the two studies include only youth in the age range that overlaps the two studies, 15- through 19-year-olds. To create age-equivalent samples, NLTS2 youth were weighted to match the age distribution of NLTS.

One other difference between the study samples that has been accommodated through analytic adjustments to enhance comparability involves the different system of disability classification in use at the time the two studies were conducted. The following adjustments have been made:

- The two NLTS categories of deaf and hard of hearing were combined to be comparable to the single NLTS2 category of hearing impairment.
- In both cohorts, students with deaf-blindness were included in the multiple impairments category because there were too few to report separately.
- Because the categories of autism and traumatic brain injury were not in use in 1987, NLTS2 students with autism or traumatic brain injury were included in other categories, using descriptions of the primary disability provided by parents. If parents said the primary disability of these students was autism or traumatic brain injury, with no other information provided, students were included in the other health impairment category, where they most likely would have been classified in 1987. If more than one disability, in addition to autism or traumatic brain injury, was mentioned by parents, students were included in the multiple impairments category. This distribution mirrors the fairly broad dispersion of NLTS students known to have autism or traumatic brain injuries.

Estimating Standard Errors

Each estimate reported in the data tables is accompanied by a standard error. A standard error acknowledges that any population estimate that is calculated from a sample will only approximate the true value for the population. The true population value will fall within the range demarcated by the estimate, plus or minus the standard error 95% of the time. For example, if the cohort 2 estimate for the current employment rate of youth out of school up to 2 years is 49%, with a standard error of 4.0 (as reported in Exhibit 5-1), one can be 95% confident that the true current employment rate for the population is between 41% and 57%.

Because the NLTS and NLTS2 samples are both stratified and clustered, calculating standard errors by formula is not straightforward. Standard errors for means and proportions were estimated using pseudo-replication, a procedure that is widely used by the U.S. Census Bureau and other federal agencies involved in fielding complex surveys. To that end, a set of weights was developed for each of 32 balanced half-replicate subsamples. Each half-replicate involved selecting half of the total set of LEAs that provided contact information using a partial factorial balanced design (resulting in about half of the LEAs being selected within each stratum) and then weighting that half to represent the entire universe. The half-replicates were used to estimate the variance of a sample mean by: (1) calculating the mean of the variable of interest on the full sample and each half-sample using the appropriate weights; (2) calculating the squares of the deviations of the half-sample estimated from the full sample estimate; and (3) adding the squared deviations and dividing by (n-1), where n is the number of half-replicates.

Although the procedure of pseudo-replication is less unwieldy than the development of formulas for calculating standard errors, it is not easily implemented using the Statistical Analysis System (SAS), the analysis program used for NLTS and NLTS2, and it is computationally expensive. Experience has demonstrated that it is possible to develop straightforward estimates of standard errors using the effective sample size.

When respondents are independent and identically distributed, the effective sample size for a weighted sample of N respondents can be approximated as

$$Neff = N \times (E^2[W] / (E^2[W] + V[W]$$

where Neff is the effective sample size, $E^2[W]$ is the square of the arithmetic average of the weights and V[W] is the variance of the weights. For a variable X, the standard error of estimate can typically be approximated by sqrt(V[X]/Neff), where V[X] is the weighted variance of X.

Respondents are not independent of each other because they are clustered in LEAs, and the intracluster correlation is not zero. However, because the intracluster correlation traditionally has been quite small, the formula for the effective sample size shown above has worked well. To be conservative, however, the initial estimate was multiplied by a "safety factor" to assure that the standard error of estimate was not underestimated.

To determine the adequacy of fit of the variance estimate based on the effective sample size and to estimate the required safety factor, 24 questions with 95 categorical and 2 continuous responses were selected. Standard errors of estimates for each response category and the mean response to each question were calculated for each disability group using both pseudo-replication and the formula involving effective sample size. A safety factor of 1.25 resulted in the effective sample size standard error estimate underestimating the pseudo-replicate standard error estimate for 92% of the categorical responses and 89% of the mean responses. Because the pseudo-replicate estimates of standard error are themselves estimates of the true standard error, and are therefore subject to sampling variability, this can be considered an adequate margin of safety.

Unweighted and Weighted Sample Sizes

As indicated above, standard errors accompany all estimates reported in the data tables. How close an estimate comes to a true population value is influenced by the size of the sample on which the estimate is based. Larger samples yield estimates with smaller standard errors, indicating that those estimates are closer to true population values than estimates with larger standard errors based on smaller samples.

The actual, or "unweighted." sample sizes for each variable reported in the data tables are included in Appendix B. However, some readers may be interested in determining the number of youth in the nation represented by a particular estimate (e.g., if 49% of youth in cohort 2 were employed currently, how many youth in the country were employed?). A first step in determining these "weighted" sample sizes involves multiplying the percentage estimate by the actual number of youth in the nation represented by that estimate (see example below). However, 95% of the time, the true population value is likely to diverge from that estimate by as much as the amount of the standard error. Therefore, it is more appropriate to use the standard error to calculate a range in the number of youth represented by an estimate, rather than relying on the single value resulting from multiplying the estimate by the size of the population it represents.

Consider the example depicted in Exhibit A-5. NLTS2 findings indicate that 48.9% of cohort 2 youth were currently employed (see Exhibit 5-1). The standard error accompanying that estimate is 4.0, indicating that the true current employment rate for the population is likely to fall between 44.9% and 52.9%. Cohort 2 represents a total of 1,455,505 youth out of school up to 2 years. Multiplying the percentages by this population size yields a single-point estimate of an estimate of 711,742 and a range of 653,522 to 769,962, within which the actual population size will fall, with 95% confidence.

Exhibit A-5
EXAMPLE OF CALCULATING WEIGHTED SAMPLE SIZES

A	B	C	D	E	F
Percentage Estimate	Standard Error	Range around Estimate (Column A Plus or Minus Column B)	Population Size	Single-point Weighted Population Affected (Column A x Column D)	Range in Weighted Population Affected (Column C x Column D)
48.9	4.0	44.9 to 52.9	1,455,505	711,742	653,522 to 769,962

Because percentage estimates are provided not only for the full sample of youth with disabilities in each cohort, but also for youth who differ in primary disability category, gender, household income, and race/ethnicity, readers must have the actual population size for each of these subgroups to calculate weighted sample sizes for some estimates. These population sizes are presented in Exhibit A-6.

Exhibit A-6
POPULATION SIZES OF GROUPS REPRESENTED BY NLTS AND NLTS2

Groups	Cohort 1	Cohort 2
All youth with disabilities	747,442	1,455,505
Disability category:		
Learning disability	447,839	729,881
Speech/language impairment	27,011	33,439
Mental retardation	139,827	149,400
Emotional disturbance	94,882	139,019
Hearing impairment	81,40	15,350
Visual impairment	3,852	5,794
Orthopedic impairment	7,341	14,061
Other health impairment	8,243	60,168
Multiple disabilities	11,217	24,839
Gender		
Boys	512,745	798,685
Girls	234,697	386,484
Household income		
Lowest	261,829	413,624
Middle	241,947	359,936
Highest	243,591	411,609
Race/ethnicity		
White	485,015	686,094
African-American	175,275	215,464
Hispanic	64,853	207,760

Calculating Significance Levels

In general, references in the text of the report to differences between groups highlight only differences that are statistically significant with at least 95% confidence (denoted as $p<.05$). Beyond the differences highlighted in the text, readers may want to compare percentages or means for specific subgroups to determine, for example, whether the difference in the percentage of students who are male between students with learning disabilities and those with hearing impairments is greater than would be expected to occur by chance. To calculate whether the difference between percentages is statistically significant, the squared difference between the two percentages of interest is divided by the sum of the two squared standard errors. If this product is larger than 3.84, the difference is statistically significant at the .05 level (i.e., it would occur by chance fewer than 5 times in 100). Presented as a formula, a difference in percentages is statistically significant at the .05 level if:

$$\frac{(P_1P_2)^2}{SE_1^2 + SE_2^2} > 1.96^2$$

where P_1 and SE_1 are the first percentage and its standard error, and P_2 and SE_2 are the second percentage and its standard error. If the product of this calculation is 6.63 to 10.79, the significance level is .01; products of 10.8 or greater are significant at the .001 level.

Measurement Issues

The chapters in this report include information on variables that were included in both NLTS and NLTS2. If there were differences between the studies in how a particular variable was defined, those differences are highlighted in the discussion of findings related to that variable. However, several general points about measures are used repeatedly in analyses that should be clear to readers as they consider the findings reported here.

Categorizing students by primary disability. Information about the nature of students' disabilities came from rosters of all students in the study age ranges who were receiving special education in the sample school years under the auspices of participating LEAs and state-supported special schools. For analysis purposes, students in both studies were assigned to a disability category on the basis of the primary disability designated by the student's school or district. Although there are federal guidelines for making category assignments criteria, methods for assigning students to categories vary from state and to state and even between districts within states, with the potential for substantial variation in the nature and severity of disabilities included in categories (see for example, MacMillan & Siperstein, 2002). Therefore, data should not be interpreted as describing students who truly had a particular disability, but rather as describing students who were categorized as having that primary disability by their school or district. Hence, descriptive data are nationally generalizable to youth out of school up to 2 years who were classified as having a particular primary disability in the school year in which they were selected for the NLTS or NLTS2 sample.

Demographic characteristics. Findings in this report are provided for youth who differ in age, gender, household income, and race/ethnicity. For the large majority of youth, age was determined from data provided by students' schools or districts. For youth for whom age information was not provided by schools or districts, birth date or age was taken from the parent interview/survey. For NLTS, gender and race/ethnicity also were obtained from parents, whereas these data were requested from and supplied by many school districts on student rosters. Classifying the income of students' households relied exclusively on information provided during the parent interview/survey. When variations in NLTS and NLTS2 variables between income groups are described, designations of lower, medium, and higher are used. These were constructed by dividing the income distribution of each study into approximate thirds. Thus, the categories indicate income relative to other youth in the study, not to a fixed income amount.

Households in poverty. A dichotomous variable indicating that a student's household was in poverty was constructed using parents' reports of household income and household size and federal poverty thresholds for 1987 and 2001 (U.S. Census Bureau, 2001). These thresholds indicate the income level for specific sizes of households, below which the household is considered in poverty. Because NLTS and NLTS2 respondents reported household income in categories (e.g., $25,000 to $29,999) rather than specific dollar amounts, estimates of poverty status were calculated by assigning each household to the mean value of the category of income reported by the parent and comparing that value to the household's size to determine poverty status.

Comparisons with the general population of students. In cases in which survey data for the general population of youth are publicly available (e.g., the National Household Education Survey), data have been abstracted from those datasets for out-of-school youth who match in age the 15- through 19-year-olds included in the comparison of NLTS and NLTS2. However, many of the comparisons have been made using published data, particularly for NLTS. For many of these comparisons, differences in samples (e.g., ages of students) or measurement (e.g., question wording on surveys) reduce the direct comparability of data for youth with disabilities and data for youth in the general population. When these limitations affect the comparisons, they are pointed out in the text and the implications for the comparisons are noted.

Reporting statistics. Statistics are not reported for groups with fewer than 35 members. Statistics with a decimal of .5 are rounded to the nearest whole even number.

APPENDIX A REFERENCES

Employment Policies Institute. (2002). *Measuring poverty in America: Science or politics.* Available at http://www.epionline.org/report_poverty_04-2002.pdf

Javitz, H. S., & Wagner, M. (1990). *The National Longitudinal Transition Study of Special Education Students. Report on sample design and limitations, wave 1 (1987).* Menlo Park, CA: SRI International.

MacMillan, D. L., & Siperstein, G.N. (2002). Learning disabilities as operationally defined by schools. In R. Bradley, L. Danielson, & D. P. Hallahan. *Identification of learning disabilities. Research to practice*. Mahwah, NJ: Lawrence Erlbaum Associates.

Office of Special Education Programs (OSEP). (2001). *Table AD1. Number of students age 14 and older exiting special education during the 1999-2000 school year*. Available at http://www.ideadata.org/tables24th/ar_ad1.htm

Wagner, M., Newman, L., & Shaver, D. (1989). *The National Longitudinal Transition Study of Special Education Students. Report on procedures for the first wave of data collection (1987)*. Menlo Park, CA: SRI International.

U.S. Census Bureau. (2001). *The 2001 HHS poverty guidelines.* Available at http://aspe.hhs.gov/poverty/01poverty.htm