

Appendix C

Data Sources

Sources and Comparability of Data

The information in this report was obtained from many sources, including federal and state agencies, private research organizations, and professional associations. The data were collected by many methods, including surveys of a universe (such as all colleges) or of a sample, and compilations of administrative records. Care should be used when comparing data from different sources. Differences in procedures, such as timing, phrasing of questions, and interviewer training, mean that the results from the different sources are not strictly comparable. More extensive documentation of one survey's procedures than of another's does not imply more problems with the data, only that more information is available on the survey.

Accuracy of Data

The accuracy of any statistic is determined by the joint effects of “sampling” and “nonsampling” errors. Estimates based on a sample will differ from the figures that would have been obtained if a complete census had been taken using the same survey instruments, instructions, and procedures. Besides sampling errors, both of the surveys, universe and sample, are subject to errors of design, reporting, and processing, and errors due to nonresponse. To the extent possible, these nonsampling errors are kept to a minimum by methods built into the survey procedures. In general, however, the effects of nonsampling errors are more difficult to gauge than those produced by sampling variability.

Sampling Errors

The standard error is the primary measure of sampling variability. It provides a specific range—with a stated confidence—within which a given estimate would lie if a complete census had been conducted. The chances that a complete census would differ from the sample by less than the standard error are about 68 out of 100. The chances that the difference would be less than 1.65 times the standard error are about 90 out of 100. The

chances that the difference would be less than 1.96 times the standard error are about 95 out of 100. The chances that it would be less than 2.58 times as large are about 99 out of 100.

The standard error can help assess how valid a comparison between two estimates might be. The standard error of a difference between two sample estimates that are uncorrelated is approximately equal to the square root of the sum of the squared standard errors of the estimates. The standard error (se) of the difference between sample estimate “a” and sample estimate “b” is

$$se_{a-b} = (se_a^2 + se_b^2)^{1/2}$$

Note that most of the standard errors in subsequent sections and in the original documents are approximations. That is, to derive estimates of standard errors that would be applicable to a wide variety of items and could be prepared at a moderate cost, a number of approximations were required. As a result, most of the standard errors presented provide a general order of magnitude rather than the exact standard error for any specific item.

Nonsampling Errors

Both universe and sample surveys are subject to nonsampling errors. Nonsampling errors are of two kinds—random and nonrandom. Random nonsampling errors may arise when respondents or interviewers interpret questions differently, when respondents must estimate values, or when coders, keyers, and other processors handle answers differently. Nonrandom nonsampling errors result from total nonresponse (no usable data obtained for a sampled unit), partial or item nonresponse (only a portion of a response may be usable), inability or unwillingness on the part of respondents to provide information, difficulty interpreting questions, mistakes in recording or keying data, errors of collection or processing, and overcoverage or undercoverage of the target universe. Random nonresponse errors usually, but not always, result in an understatement of sampling errors and thus an overstatement of the precision of survey estimates. Because estimating the magnitude of nonsampling errors would require special experiments or access to independent data, these magnitudes are seldom available.

To compensate for suspected nonrandom errors, adjustments of the sample estimates are often made. For example, adjustments are frequently made for nonresponse, both total and partial. Imputations are usually made separately within various groups of sample members that have similar survey characteristics. Imputation for item nonresponse is usually made by substituting for a missing item the response to that item of a respondent having characteristics similar to those of the respondent.

Although the magnitude of nonsampling errors in the data used in *Projections of Education Statistics* is frequently unknown, idiosyncrasies that have been identified are noted on the appropriate tables.

Federal Agency Sources

National Center for Education Statistics (NCES)

Common Core of Data

NCES uses the Common Core of Data (CCD) survey to acquire and maintain statistical data from each of the 50 states, the District of Columbia, the Bureau of Indian Affairs, Department of Defense Dependents' Schools (overseas), and the outlying areas. Information about staff and students is collected annually at the school, local education agency or school district (LEA), and state levels. Information about revenues and expenditures is also collected at the state and LEA levels.

Data are collected for a particular school year (July 1 through June 30) via survey instruments sent to the state education agencies during the school year. States have 1 year in which to modify the data originally submitted.

Since the CCD is a universe survey, the CCD information presented in this edition of the *Projections of Education Statistics* is not subject to sampling errors. However, nonsampling errors could come from two sources—nonreturn and inaccurate reporting. Almost all of the states submit the six CCD survey instruments each year, but submissions are sometimes incomplete or too late for publication.

Understandably, when 58 education agencies compile and submit data for approximately 95,000 public schools and 17,000 local school districts, misreporting can occur. Typically, this results from varying interpretations of NCES definitions and differing recordkeeping systems. NCES attempts to minimize these errors by working closely with the state education agencies through the National Forum on Education Statistics.

The state education agencies report data to NCES from data collected and edited in their regular reporting cycles. NCES encourages the agencies to incorporate into their own survey systems the NCES items they do not already collect so that those items will also be available for the subsequent CCD survey. Over time, this has meant fewer missing data cells in each state's response, reducing the need to impute data.

NCES subjects data from the education agencies to a comprehensive edit. Where data are determined to be inconsistent, missing, or out of range, NCES contacts the education agencies for verification. NCES-prepared state summary forms are returned to the state education agencies for verification. States are also given an opportunity to revise their state-level aggregates from the previous survey cycle.

Further information on the CCD may be obtained from

John Sietsema
Elementary/Secondary Cooperative System
and Institutional Studies Division (ESCSISD)
National Center for Education Statistics
1990 K Street NW
Washington, DC 20006
John.Sietsema@ed.gov
<http://nces.ed.gov/ccd/>

Private School Universe Survey

The purposes of Private School Survey (PSS) data collection activities are to build an accurate and complete list of private schools to serve as a sampling frame for NCES sample surveys of private schools, and to report data on the total number of private schools, teachers, and students in the survey universe. The PSS is conducted every 2 years, with collections in the 1989–90, 1991–92, 1993–94, 1995–96, 1997–98, 1999–2000, 2001–02, and 2003–04 school years.

The PSS produces data similar to that of the CCD for public schools and can be used for public-private comparisons. The data are useful for a variety of policy and research-relevant issues, such as the growth of religiously affiliated schools, the number of private high school graduates, the length of the school year for various private schools, and the number of private school students and teachers.

The target population for the universe survey consists of all private schools in the United States that meet NCES criteria of a school (e.g., a private school is an institution that provides instruction for any of

grades K through 12, has one or more teachers to give instruction, is not administered by a public agency, and is not operated in a private home). The survey universe is composed of schools identified from a variety of sources. The main source is a list frame, initially developed for the 1989–90 PSS. The list is updated regularly, matching it with lists provided by nationwide private school associations, state departments of education, and other national guides and sources that list private schools. The other source is an area frame search in approximately 120 geographic areas, conducted by the Census Bureau.

Further information on the PSS may be obtained from

Steve Broughman
Elementary/Secondary Sample Survey
Studies program (ESLSD)
National Center for Education Statistics
1990 K Street NW
Washington, DC 20006
Stephen.Broughman@ed.gov
<http://nces.ed.gov/surveys/pss/>

Integrated Postsecondary Education Data System

The Integrated Postsecondary Education Data System (IPEDS) surveys approximately 6,500 postsecondary institutions, including universities and colleges, as well as institutions offering technical and vocational education beyond the high school level. This survey, which began in 1986, replaced the Higher Education General Information Survey (HEGIS).

IPEDS consists of eight integrated components that obtain information on who provides postsecondary education (institutions), who participates in it and completes it (students), what programs are offered and what programs are completed, and both the human and financial resources involved in the provision of institutionally-based postsecondary education. Until 2000 these components included: Institutional Characteristics, Fall Enrollment, Completions, Salaries, Finance, and Fall Staff. Data are collected in the fall for institutional characteristics and completions; in the winter for employees by assigned position (EAP), salaries and fall staff; and in spring for enrollment, student financial aid, finances, and graduation rates.

The degree-granting institutions portion of this survey is a census of colleges awarding associate's or higher degrees, that were eligible to participate in Title IV financial aid programs. Prior to 1993, data from the technical and vocational institutions were collected through a sample

survey. Beginning in 1993, all data were gathered in a census of all postsecondary institutions. The IPEDS tabulations developed for this edition of *Projections of Education Statistics* are based on lists of all institutions and are not subject to sampling errors.

The definition of institutions generally thought of as offering college and university education has been changed in recent years. The old standard for higher education institutions included those institutions that had courses that led to an associate degree or higher, or were accepted for credit towards those degrees. The higher education institutions were accredited by an agency or association that was recognized by the U.S. Department of Education or recognized directly by the Secretary of Education. The current category includes institutions that award associate or higher level degrees that are eligible to participate in Title IV federal financial aid programs. The impact of this change has generally not been large. For example, tables on faculty salaries and benefits were only affected to a very small extent. Also, degrees awarded at the bachelor's level or higher were not heavily affected. Most of the data on public 4-year colleges have been affected only to a minimal extent. The impact on enrollment in public 2-year colleges was noticeable in certain states, but relatively small at the national level. The largest impact has been on private 2-year college enrollment. Overall, enrollment for all institutions was about one-half of a percent higher for degree-granting institutions compared to the total for higher education institutions.

Prior to the establishment of IPEDS in 1986, HEGIS acquired and maintained statistical data on the characteristics and operations of institutions of higher education. Implemented in 1966, HEGIS was an annual universe survey of institutions accredited at the college level by an agency recognized by the Secretary of the U.S. Department of Education. These institutions were listed in the NCES publication *Education Directory, Colleges and Universities*.

HEGIS surveys solicited information concerning institutional characteristics, faculty salaries, finances, enrollment, and degrees. Since these surveys were distributed to all higher education institutions, the data presented are not subject to sampling error. However, they are subject to nonsampling error, the sources of which varied with the survey instrument. Information concerning the nonsampling error of the HEGIS enrollment and degrees surveys can be obtained from the HEGIS Post Survey Validation Study conducted in 1979.

Elise Miller
Postsecondary Institutional Studies Program (PSD)
National Center for Education Statistics
1990 K Street NW
Washington, DC 20006
Elise.Miller@ed.gov
<http://nces.ed.gov/ipeds/>

Fall (Institutional Characteristics) This survey collects basic information necessary to classify the institutions, including control, level, and kinds of programs, and information on tuition, fees, and room and board charges. Beginning in 2000, the survey collected institutional pricing data from institutions with first-time, full-time, degree/certificate-seeking undergraduate students. Unduplicated full-year enrollment counts and instructional activity are now collected on the Fall Enrollment survey. The overall response rate was 100.0 percent for Title IV degree-granting institutions in 2003.

Further information may be obtained from

Patricia Brown
Postsecondary Institutional Studies Program (PSD)
National Center for Education Statistics
1990 K Street NW
Washington, DC 20006
Patricia.Brown@ed.gov
<http://nces.ed.gov/ipeds/>

Winter/Spring (Fall Enrollment) This survey has been part of the HEGIS and IPEDS series since 1966. The enrollment survey response rate is relatively high. Beginning in 2000, the data collection method was web-based, replacing the paper survey forms that had been used in past years. In 2004 the overall response rate was 100.0 percent for degree-granting institutions and the 2003 response rate was 99.6 percent. Imputation methods and response bias analysis for the 2003-04 survey are discussed in *Enrollment in Postsecondary Institutions, Fall 2003; Graduation Rates 1997 and 2000 Cohorts; and Financial Statistics, Fiscal Year 2003* (NCES 2005-177). Imputation methods and response bias analysis for the 2002-03 survey are discussed in *Enrollment in Postsecondary Institutions, Fall 2002 and Financial Statistics, Fiscal Year 2002* (NCES 2005-168).

Beginning with fall 1986, the survey system was redesigned with the introduction of IPEDS (see above). The survey allows (in alternating years) for the collection

of age and residence data. In 2000, the enrollment survey collected the instructional activity and unduplicated headcount data, which are needed to compute a standardized, full-time equivalent (FTE) enrollment statistic for the entire academic year.

Integrated Postsecondary Education Data System Data Quality Study, Methodology Report, July 2005 (NCES 2005-175) showed that public institutions made the majority of changes to enrollment data during the 2004 revision period. The majority of changes were made to unduplicated headcount data with the net differences between the original data and the revised data at about 1 percent. Part-time students in general and enrollment in private not-for-profit institutions were often underestimated. The fewest changes by institutions were to CIP code data. More institutions provided enrollment data to IPEDS than to Thomson Peterson. A fairly high percentage of institutions that provided data to both provided the same data, and among those that did not, the difference in magnitude was less than 10 percent.

Further information may be obtained from

Cathy Statham
Postsecondary Institutional Studies Program (PSD)
National Center for Education Statistics
1990 K Street NW
Washington, DC 20006
Cathy.Statham@ed.gov
<http://nces.ed.gov/ipeds/>

Fall (Completions) This survey was part of the HEGIS series throughout its existence. However, the degree classification taxonomy was revised in 1970–71, 1982–83, 1991–92, and 2002–03. Collection of degree data has been maintained through the IPEDS system.

The nonresponse rate did not appear to be a significant source of nonsampling error for this survey. The return rate over the years has been high, with the degree-granting institutions response rate for the 2003-04 survey at 99.8 percent. The overall response rate for the non-degree granting institutions was 99.6 percent in 2003-04. Because of the high return rate for the degree-granting institutions, nonsampling error caused by imputation is also minimal. Imputation methods and response bias analysis for the 2003-04 survey are discussed in *Postsecondary Institutions in the United States: Fall 2004 and Degrees and Other Awards Conferred: 2003-04* (NCES 2005-154).

The Integrated Postsecondary Education Data System Data Quality Study, Methodology Report, July 2005 (NCES 2005-175) indicated that most Title IV institutions supplying revised data on completions were able to supply missing data for the prior year. The small differences between imputed data for the prior year and the revised actual data supplied by the institution indicated that the imputed values produced by NCES were acceptable.

Further information on IPEDS Completions surveys may be obtained from

Andrew Mary
 Postsecondary Institutional Studies Program (PSD)
 National Center for Education Statistics
 1990 K Street NW
 Washington, DC 20006
Andrew.Mary@ed.gov
<http://nces.ed.gov/ipeds/>

Census Bureau

Current Population Survey

Prior to July 2001, estimates of school enrollment rates, as well as social and economic characteristics of students, were based on data collected in the Census Bureau's monthly household survey of about 50,000 dwelling units. Beginning in July 2001, this sample was expanded to 60,000 dwelling units. The monthly Current Population Survey (CPS) sample consists of 754 areas comprising 2,007 geographic areas, independent cities, and minor civil divisions throughout the 50 states and the District of Columbia. The samples are initially selected based on the decennial census files and are periodically updated to reflect new housing construction.

The monthly CPS deals primarily with labor force data for the civilian noninstitutional population (i.e., excluding military personnel and their families living on post and inmates of institutions). In addition, in October of each year, supplemental questions are asked about highest grade completed, level and grade of current enrollment, attendance status, number and type of courses, degree or certificate objective, and type of organization offering instruction for each member of the household. In March of each year, supplemental questions on income are asked. The responses to these questions are combined with answers to two questions on educational attainment: highest grade of school ever attended, and whether that grade was completed.

The estimation procedure employed for monthly CPS data involves inflating weighted sample results to independent estimates of characteristics of the civilian noninstitutional population in the United States by age, sex, and race. These independent estimates are based on statistics from decennial censuses; statistics on births, deaths, immigration, and emigration; and statistics on the population in the armed services. Generalized standard error tables are provided in the *Current Population Reports*. The data are subject to both nonsampling and sampling errors.

Caution should also be used when comparing data between Census years. With the release of the January 2003 CPS data, population controls that reflect the results of Census 2000 were used in the monthly CPS estimation process. The new controls increased the size of the civilian noninstitutional population by about 3.5 million in May 2002. This adjustment usually occurs 3 to 4 years after the census, and, if the adjustment is substantial, historical data will be revised. Data from January 2000 through December 2002 were revised to reflect these new controls. Over and above these revisions, the U.S. Census Bureau introduced another large upward adjustment to the controls as part of its annual update of population estimates for 2003. The prior change in population controls occurred in March 1993, where data after this date were based on the 1990 census-based population controls and data before this date were based on 1980 or earlier census based population controls. This change in population controls between 1980-based and 1990-based had relatively little impact on summary measures, such as means, medians, and percentage distributions. It does, however, have a significant impact on levels. For example, use of 1990-based population controls resulted in about a 1 percent increase in the civilian noninstitutional population and in the number of families and households. Thus, estimates of levels for data collected in 1994 and later years differed from those for earlier years by more than what could be attributed to actual changes in the population. These differences could be disproportionately greater for certain subpopulation groups than for the total population.

In addition to the changes in population controls, two other relevant changes were introduced into the CPS with the release of the January 2003 data. First, the questions on race and Hispanic origin in the CPS were modified to comply with the new standards for maintaining, collecting, and presenting Federal data on race and ethnicity for Federal statistical agencies. A major change under those standards is that respondents may select more

than one race when answering the survey. Respondents continued to be asked a separate question to determine if they are Hispanic, which is considered an ethnicity rather than a race. The ethnicity question was reworded to ask directly whether the respondent was Hispanic. Persons who report they are Hispanic also are classified separately in the race (or races) they consider themselves to be. Second, improvements were introduced to both the second stage and composite weighting procedures. These changes adapt the weighting procedures to the new race/ethnic classification system and enhance the stability over time of national and state/substate labor force estimates for demographic groups. These two changes, in addition to the change in population controls discussed above, benchmark the CPS data to the results of Census 2000, improve the estimation procedures, and ensure that the data series produced from the survey reflect the evolving composition of the U.S. population.

Further information on CPS may be obtained from

Education and Social Stratification Branch
 Population Division
 Census Bureau
 U.S. Department of Commerce
 Washington, DC 20233
<http://www.bls.census.gov/cps/cpsmain.htm>

School Enrollment Each October, the Current Population Survey (CPS) includes supplemental questions on the enrollment status of the population 3 years old and over, in addition to the monthly basic survey on labor force participation. Prior to 2001, the October supplement consisted of approximately 47,000 interviewed households. Beginning with the October 2001 supplement, the sample was expanded by 9,000 to a total of approximately 56,000 interviewed households. The main sources of nonsampling variability in the responses to the supplement are those inherent in the survey instrument. The question of current enrollment may not be answered accurately for various reasons. Some respondents may not know current grade information for every student in the household, a problem especially prevalent for households with members in college or in nursery school. Confusion over college credits or hours taken by a student may make it difficult to determine the year in which the student is enrolled. Problems may occur with the definition of nursery school (a group or class organized to provide educational experiences for children), where respondents' interpretations of "educational experiences" vary.

For the October 2001 basic CPS, the nonresponse rate was 6.7 percent, and for the school enrollment supplement, the nonresponse rate was an additional 3.6 percent, for a total supplement nonresponse rate of 10.1 percent.

Further information on CPS methodology may be obtained from

<http://www.bls.census.gov/cps/cpsmain.htm>

Further information on CPS "School Enrollment" may be obtained from

Education and Social Stratification Branch
 Census Bureau
 U.S. Department of Commerce
 Washington, DC 20233
<http://www.census.gov/population/www/socdemo/school.html>

State Population Projections These state population projections were prepared using a cohort-component method by which each component of population change—births, deaths, state-to-state migration flows, international in-migration, and international out-migration—was projected separately for each birth cohort by sex, race, and Hispanic origin. The basic framework was the same as in past Census Bureau projections.

Detailed components necessary to create the projections were obtained from vital statistics, administrative records, census data, and national projections.

The cohort-component method is based on the traditional demographic accounting system:

$$P_1 = P_0 + B - D + DIM - DOM + IIM - IOM$$

where:

- P_1 = population at the end of the period
- P_0 = population at the beginning of the period
- B = births during the period
- D = deaths during the period
- DIM = domestic in-migration during the period
- DOM = domestic out-migration during the period
- IIM = international in-migration during the period
- IOM = international out-migration during the period

To generate population projections with this model, the Census Bureau created separate datasets for each of these components. In general, the assumptions concerning the future levels of fertility, mortality, and international migration are consistent with the assumptions developed for the national population projections of the Census Bureau.

Once the data for each component were developed, it was a relatively straightforward process to apply the cohort-component method and produce the projections. For each projection year, the base population for each state was disaggregated into eight race and Hispanic categories (non-Hispanic White; non-Hispanic Black; non-Hispanic American Indian, Eskimo, and Aleut; non-Hispanic Asian and Pacific Islander; Hispanic White; Hispanic Black; Hispanic American Indian, Eskimo, and Aleut; and Hispanic Asian and Pacific Islander), by sex, and single year of age (ages 0 to 85+). The next step was to survive each age-sex-race-ethnic group forward 1 year using the pertinent survival rate. The internal redistribution of the population was accomplished by applying the appropriate state-to-state migration rates to the survived population in each state. The projected out-migrants were subtracted from the state of origin and added to the state of destination (as in-migrants). Next, the appropriate number of immigrants from abroad was added to each group. The population under age 1 was created by applying the appropriate age-race-ethnic-specific birth rates to females of childbearing age. The number of births by sex and race/ethnicity were survived forward and exposed to the appropriate migration rate to yield the population under age 1. The final results of the projection process were adjusted to be consistent with the national population projections by single years of age, sex, race, and Hispanic origin. The entire process was then repeated for each year of the projection.

More information is available in the Census Bureau Population Paper Listing 47 (PPL-47) and Current Population Report P25-1131. These reports may be obtained from

Statistical Information Staff
Census Bureau
U.S. Department of Commerce
Washington, DC 20233
(301) 763-3030
<http://www.census.gov>

Other Sources

National Education Association

Estimates of School Statistics

The National Education Association (NEA) reports enrollment, teacher, revenue, and expenditure data in its annual publication *Estimates of School Statistics*. Each year, NEA prepares regression-based estimates of financial and other education statistics and submits them to the states for verification. Generally, about 30 states adjust these estimates based on their own data. These preliminary data are published by NEA along with revised data from previous years. States are asked to revise previously submitted data as final figures become available. The most recent publication contains all changes reported to the NEA.

Additional information is available from

National Education Association—Research
1201 16th Street NW
Washington, DC 20036
<http://www.nea.org>

Global Insight, Inc.

Global Insight, Inc. provides an information system that includes: databases of economic and financial information; simulation and planning models; regular publications and special studies; data retrieval and management systems; and access to experts on economic, financial, industrial, and market activities. One service is the Global Insight Model of the U.S. Economy, which contains annual projections of U.S. economic and financial conditions, including forecasts for the federal government, incomes, population, prices and wages, and state and local governments, over a long-term (10- to 25-year) forecast period.

Additional information is available from

Global Insight, Inc.
1000 Winter Street
Suite 4300N
Waltham, MA 02451-124
<http://www.globalinsight.com/>

