

## Glossary of Specialized and Technical Terms Used in the ESCAP Report and Supporting Documents<sup>1</sup>

**Accuracy and Coverage Evaluation (A.C.E.):** The name given to the program to measure and correct for net coverage errors in Census 2000. The A.C.E. includes selecting a sample of block clusters, address listing, housing unit matching, interviewing, person matching, follow-up, post-stratum estimation, synthetic estimation, controlled rounding and record imputation. Both the Population Sample and the Enumeration Sample are part of the A.C.E.

**A.C.E.:** See Accuracy and Coverage Evaluation

**A.C.E. Universe:** The household population, excluding areas in remote Alaska enumerated with special methods. Specifically excludes group quarters and service based (often called “homeless”) populations.

**Alternative DA:** Term given to demographic analysis estimates (or scenarios) produced in February, 2001 that included a doubling of the number of unauthorized immigrants who arrived during the 1990 that were included in the base DA. This revised number implied a foreign-born population roughly consistent with results from the March, 2001 Current Population Survey as reweighted using Census 2000 counts a population controls.

**Artificial Population:** Simulated population totals for small areas based on distributing the estimated net undercount using surrogate variables correlated with the actual undercount. They are used to assess the impact of the A.C.E. assumption of uniform net undercounts within poststrata for all geographic areas.

**Balancing Error:** An error in the DSE that occurs when one set of rules is used to determine which records are included in the census for determining the number of omissions and another set of rules is used to determine which records are correctly included in the census. For example, balancing error would occur if the P-sample matching process was allowed to match to a census record three miles away from its correct location, but that same census unit would be considered erroneous if sampled in the E sample.

**Bias:** The difference between the true number and the expected value of an estimate. The error that would be produced in the absence of sampling variance.

**Bias in the Synthetic Estimate:** The difference between the true population of an area and the synthetic estimate. This includes both the bias in the post-stratum estimates and the synthetic error.

---

<sup>1</sup> The definitions in this glossary are, in some cases, simplified and are provided for the general understanding of the reader. Readers requiring more precision are directed to the official definitions contained in the census documentation.

Base DA: The demographic analysis results produced before February 2001 that used assumptions about the growth of the unauthorized population based on calculations made in 1992 using data from the 1990 census.

Block Cluster: The basic sample unit for the A.C.E. A block cluster usually consists of either a census block or a group of contiguous blocks.

CCF: See Coverage Correction Factor

CAPI: See Computer Assisted Personal Interviewing

Causal Dependence: The tendency of some people to be either more or less likely to be included in the A.C.E. because they had been included in the census, or vice versa.

Census: The census includes all activities related to the Decennial Census of Population and Housing, and includes the A.C.E. However, in the context of the ESCAP reports, the term is often used to include decennial activities such as address listing and enumeration excluding the A.C.E.

Census Count: The total official population count including, specifically, correct enumerations, erroneous enumerations, whole person imputations, and late census additions.

Cluster: See block cluster

Computer Assisted Personal Interviewing (CAPI): The use of laptop computers to conduct the initial (May-August) A.C.E. Person Interviewing.

Contamination Error: Causal dependence, specifically the tendency of the people in the A.C.E. blocks to be enumerated differently in the census because of A.C.E. activities.

Content Error: Error in the characteristics that are reported for those persons or housing units that are enumerated.

Controlled Rounding: In the A.C.E. context, this is a mathematical process that converts fractional estimates of net undercount for small areas into integer (whole person) counts, and does it in such a way that the totals for larger areas and larger groups remain unchanged.

Correlation Bias: The bias in the dual system estimates because of the tendency for people enumerated in the census to be more likely to be included in the A.C.E. than those missed in the census. Correlation bias usually results in a downward bias in the DSE. Correlation bias may be due to either causal dependence or heterogeneity bias.

Coverage Error: Error in an estimate that results from (a) failure to include in a census or a survey all eligible people or units or (b) inclusion of some people or units erroneously.

Coverage Correction Factor (CCF): The multiplicative factor used to compute adjusted data for small areas. The CCF is specifically the ratio of the dual system estimate to the census count and is defined for each poststratum.

Coverage Measurement Survey: A generic term used to describe any survey used to measure the coverage error of a census. Both the A.C.E. and the 1990 PES are coverage measurement surveys.

DA: See Demographic Analysis

Data-defined: Census person records with sufficient data to be accepted for further processing. Data-defined records for Census 2000 must have at least two completed items. One item may be name, defined as at least three characters in the name field. Records that are not data-defined are whole person imputations.

Demographic Analysis (DA): A technique to estimate the size of the population using birth records, death records, immigration records, Medicare records, and the results of the current and previous censuses. Demographic analysis has been a standard Census Bureau method since 1970 to estimate the net national undercount of the Census. Demographic analysis estimates are produced for both the Black and the non-Black population.

Discrepant Enumeration: Census records that do not refer to actual people in the cluster.

Domain: See Estimation Domain.

DSE: Dual System Estimate, an estimate of the true population produced by Dual System Estimation at the poststratum level.

Dual System Estimation A technique that estimates the true population using estimates of the number of people correctly included in the census and the ratio of the number of people correctly included in the census to the total true population. The number of people is estimated via the Enumeration sample. The ratio is estimated via the Population sample.

Duplicate Enumeration: A person record that refers to a person also included elsewhere in the census. The A.C.E. measures duplicates only within the search area.

EE: See Erroneous Enumeration

EFU: See Evaluation Follow-up

Enumeration Sample (E sample): A sample of census enumerations, specifically data-defined census person records. In 2000, the sample was drawn from census

housing unit records geocoded (correctly or incorrectly) to the A.C.E. sample block clusters.

**Erroneous Enumeration:** A data-defined census record that is in some way defective. In the context of the A.C.E., the record is either not sufficiently complete to adequately define a single person (that is, lacks a valid name and has fewer than two other demographic characteristics.) or refers to a person who did not usually live within the block cluster or search area on April 1, 2000. In other contexts, a more restrictive definition may be appropriate.

**E Sample:** See Enumeration Sample

**Estimation Domain:** Seven groupings of the population based on race and Hispanic origin used in the A.C.E. to form post-strata.

**Evaluation Follow-up (EFU):** A special follow-up of a sample of A.C.E. records conducted in January and February 2001 to detect errors in the A.C.E. interviewing process, which includes the original person interview and the follow-up interviewing. The results of this sample were coded two times using different rules, and thus, although there was only one evaluation follow-up, there are references to EFU-1 and EFU-2.

**Evaluation Poststratum:** One of sixteen poststrata for which A.C.E. error components are directly measured, and thus can be used to produce sets of population estimates corrected for A.C.E. probable error. The Evaluation Poststrata are used in the Total Error Model.

**Extended Search:** Searching in the blocks adjacent to the block cluster to find additional matches or additional duplicates. The process included additional field work to locate geocoding errors in the surrounding blocks. More generally, the process of converting some P-sample non-matches to matches and some E-sample erroneous enumerations to correct enumerations because they were found in an adjacent block.

**Fictitious Enumeration:** Census records that do not refer to actual people, including cartoon characters, dogs, cats, and “people” made up by the respondent or the census taker.

**Follow-up:** In the A.C.E. context, a re-interview to help determine whether a person usually lived at the address on April 1, 2000 and to resolve other matching and coding difficulties.

**Foreign Born:** Concept used in demographic analysis. The foreign born includes resident naturalized citizens, permanent legal residents, temporary migrants, refugees, unauthorized immigrants, etc. The foreign born excludes those born-abroad to US citizens and those who moved from US territories.

Gross Coverage Error: A measure of census quality that adds together census omissions and census erroneous enumerations.

Group Quarters: Dwelling places that are not housing units. These includes both institutions (such as prisons) and other group quarters (such as college dormitories).

Heterogeneity Bias: A type of correlation bias due to the fact that different people within poststrata both have different chances of being counted in the census and different chances of being included in the A.C.E. To cause a bias, these chances must be correlated, that is most often, those most likely to be missed by the census are also most likely to be missed by the A.C.E.

Household: A group of people living in a housing unit.

Housing Unit: A housing unit is a house, an apartment, a mobile home or trailer, a group of rooms or a single room occupied as a separate living quarters, or, if vacant, intended for occupancy as a separate living quarters.

Housing Unit Duplication Operation: A special census project conducted in October 2000 aimed at removing people who were duplicated in the Census because their housing unit was duplicated on the census address files. Person records in suspected duplicate housing units were analyzed and either reinstated or deleted.

HUDO: See Housing Unit Duplication Operation

Imputation: A general statistical technique that accounts for missing data by assigning a specific value for the missing item. Imputation is used by the Census, the A.C.E. and the evaluation studies.

In-mover: A person who lives in a sample housing unit at the time of the A.C.E. interview who did not reside there on Census Day, April 1, 2000.

II: A term used for whole person imputations in the census. See Insufficient Information for Matching

Insufficient Information for Matching: Incomplete census records that are excluded from the A.C.E. matching. Two kinds of records that are considered "insufficient information for matching." First, there are whole person imputations, that is records that the regular census process considers as not data-defined. These records are denoted as II, and are not part of the E sample. In addition, there are census data-defined records which are sampled in the E sample, but which the A.C.E. process codes as "erroneous enumerations" because they do not pass the more stringent requirements for A.C.E. matching and follow-up. Most commonly, these records lack a usable or complete name.

Late Census Data: In 2000, Some census person records were set aside during the Housing Unit Duplicate Operation and then reinstated in the census. In 1990, the

- term included census enumerations conducted very late in the process (i.e. November).
- Legal Permanent Resident:** A term used in demographic analysis to denote residents with long term status in the country.
- Listing:** The process of listing addresses in an area. Both the A.C.E. and the Census conduct address listing operations.
- Long Form:** The approximately 1-in-6 sample to collect detailed characteristic information. In this context, the important items relate to citizenship and year of entry used in the demographic analysis.
- Long Form Data-Defined:** Person records from households selected for the content (“long form”) sample that are sufficiently complete to be used in the long-form estimation process.
- Loss Function Analysis:** Analysis that assess the relative accuracy of the adjusted and the unadjusted census numbers for state and local areas. The loss function analysis employs simulation methods using data from the evaluation studies on bias and variances. These analyses produce accuracy measures for levels and shares.
- Mail Return Rate:** See Return Rate
- Matching:** In the A.C.E. context, the process of searching the census records from the cluster and search area to see if someone in the P sample is included in the census. Matching is also used in many other census operations.
- Match Probability:** The probability that a P-sample record is included in the census. It equals one for clear matches and zero for clear non-matches. However, for unresolved cases, an intermediate value may be assigned.
- Matching Error Study:** An evaluation rematch of a sub-sample of A.C.E. clusters to measure matching error of the A.C.E. production matching process.
- Measurement Error Reinterview:** An evaluation study to measure data collection error using the results from the Evaluation Follow-up Study.
- Mover:** A person who changes usual residence between Census Day and the time of the A.C.E. interview. Includes both in-movers and out-movers.
- Non-mover:** A person who lived in a sample housing unit both on Census Day and at the time of the A.C.E. interview.
- Nonresponse:** A general term for any missing data caused by the respondent or interviewer. Whole household non-response in the A.C.E. is handled by a separate process from missing characteristics, enumeration status, or residence status.

Nonresponse Follow-up (NRFU): The census field operation to visit all housing units for which a questionnaire has not been returned by mail, telephone or the Internet.

Omission: A person who should have been counted in the census but was not. Within the A.C.E. context, an omission is a person who should have been included but was not correctly and completely enumerated. People who were only counted in the wrong place or counted with insufficient information for matching are treated as census omissions.

Other Counting Errors: Enumerations that are considered erroneous because the person did not usually live on April 1 in the area where he/she was enumerated. These errors include enumerations of people who died before April 1, who were born after April 1, who moved out before April 1 or who moved in after April 1.

Other Residence: When used in the context of erroneous enumerations, this means the same as "Other counting errors."

Out-mover: A person who lived in a sample housing unit on Census Day, April 1, 2000, but who no longer lived there at the time of the A.C.E. interview.

Person Follow-up (PFU): Interviews conducted in the fall of 2000 as part of the A.C.E. Interviewers sought to determine the residence status of non-matched E-sample cases and some non-matched P-sample cases, as well as resolve other issues with the matching.

Person Interviewing (PI): The initial independent A.C.E. interview, conducted in May to August 2000.

PES: See Post-Enumeration Survey

PFU: See Person Follow-up

Population Sample (P sample): An independent sample of people who should have been included in the Census. For the A.C.E., the Population Sample includes people who, on April 1, usually resided in housing units in the sample block clusters. The Population Sample also included any additional people those living in those block clusters at the time of the A.C.E. person interview, that is, people who had moved into the blocks since April 1.

Post-Enumeration Survey (PES): A survey conducted after the census used to evaluate the census, often using dual system estimation. This term is sometimes used to specify the 1990 Post-Enumeration Survey.

P Sample: See Population Sample

Post-stratification: Dividing the sample into separate estimation cells according to domains based on race and Hispanic origin, tenure, and other factors. Post-stratification is used for both in dual system estimation and synthetic estimation.

Post-stratum: An estimation cell. The A.C.E. defined 449 separate estimation cells based on race/ethnicity, Census return rate, tenure, and geography.

Return Rate: The number of census questionnaires responded to by mail, telephone, or Internet divided by the occupied housing units from which a return was expected.

Residence Rules: A series of rules that adapts the concept of usual residence on Census Day (April 1) to a specific location. The rule that college students living in the dormitory are to be counted at college is an example of a residence rule.

Residual Population: A term used in demographic analysis that includes unauthorized (undocumented) residents together with other foreign-born population groups for which no direct administrative data are available.

Reinstated People: People in housing units set aside during the Housing Unit Duplication Operation and then reinstated in the census files. Also known as Late Census Additions.

Reinstated Units: Housing units set aside during the Housing Unit Duplication Operation and then reinstated in the census files.

Revised DA: The demographic analysis results of October 2001 based on a recalibration of all components of demographic analysis. The major changes from the February demographic analyses included assumptions of improved birth registration, the use of estimates of the foreign-born population based on preliminary Census 2000 data, and an allowance for undercoverage of the foreign-born population counted in 2000.

Special Place: A building or complex that includes group quarters and may also include housing units.

Synthetic Error: The error introduced at smaller geographic areas when the assumption that the net undercount is geographically uniform within poststrata is not correct.

Synthetic Estimates: Estimates by applying the rates for a larger group or area to all the smaller groups or areas that comprise it. For example, the coverage rate measured at the national level for an age, sex, race, Hispanic origin group is assumed to apply to that group for all states, cities, and smaller areas.

Targeted Extended Search (TES): The A.C.E. operation that searched for additional matches within the blocks that surrounded the block clusters in the A.C.E. sample. The search was specifically targeted to find whole households missed in the sample cluster but included in the surrounding block. Targeted extended search also included verifying whether whole households of people incorrectly included by the census within the sample cluster were in fact located in the surrounding blocks. The process also includes extended search for duplicates.



**Target Populations:** A set of state or local population estimates corrected for the bias in the A.C.E. Since the A.C.E. biases are estimated with uncertainty, a distribution of target populations is used to assess the relative accuracy of the adjusted and unadjusted population data.

**Temporary Migrant:** A person legally resident within the U.S. but whose status does not allow for permanent residence or eventual naturalization.

**Tenure:** Whether a person lives in a housing unit that is owned or being bought by himself or another resident; the two categories are (a) owner and (b) renter or other non-owner.

**TES** See Targeted Extended Search

**TES 2,3:** Evaluation studies conducted in the summer of 2001 to assess the impact of census and A.C.E. geocoding errors on the DSE and on the apparent lack of balance.

**Total Error Model:** A statistical model that attempts to estimate biases (systematic errors) in the A.C.E. estimates. Since many of the bias estimates are based on a follow-up sample survey, the Total Error Model provides both estimates of bias and statistical measures of uncertainty (variances) for the bias estimates.

**Unauthorized Population:** Foreign-born population residing in the United States without legal status.

**Undercount:** The difference between the true population and the official census count. Technically, undercount includes both coverage and content errors. For large populations with limited content errors, undercount will be the same as undercoverage.

**Undercoverage:** The difference between the true population and the number of population members included in the census, regardless of whether they were properly classified.

**Unduplication:** The process of removing duplicate enumerations or housing units from the census files.

**Usual Residence:** Usual residence is defined as the place where the person lives and sleeps most of the time. Applying the usual residence concept to real living situations means that people will not always be counted at the place where they happen to be staying on Census Day. People who live at more than one residence during the week, month, or year are to be counted at the place where they live most of the time.

**Variance:** A measure of variability in an estimate. Sampling variance is a measure of the variability introduced because only a random part, rather than all, of the population is selected for interview or inclusion.

Whole Person Imputation: A census record created by entirely replicating the characteristics of another census record. Whole person imputation records need to be created when the original census record lacks sufficient information to be considered “Data-defined.” Whole person imputations are subtracted from the census count when computing the DSE but are included when calculating the net undercount.