

**1995 COMMERCIAL BUILDING ENERGY CONSUMPTION SURVEY
SAMPLE SELECTION PROCEDURES MANUAL**

Volume 1 of 2

Prepared for:

**UNITED STATES DEPARTMENT OF ENERGY
ENERGY INFORMATION ADMINISTRATION
OFFICE OF ENERGY MARKETS AND END USE
ENERGY END USE AND INTEGRATED STATISTICS DIVISION
RESIDENTIAL AND COMMERCIAL BRANCH
WASHINGTON, DC 20585**

Prepared by:

**RESPONSE ANALYSIS CORPORATION
377 WALL STREET
PRINCETON, NJ 08542**

March 1996

Contract Number: DE-AC01-92E121948

CHAPTER 1: INTRODUCTION AND OVERVIEW	1-1
1.1 CB ECS SAMPLE DESIGN CONSIDERATIONS	1-1
1.2 DESCRIPTION OF THE UNIVERSE	1-2
1.3 AREA FRAME SAMPLING PROCEDURES	1-4
1.3.1 Selecting Area Frame Segments	1-4
1.3.2 Developing Area Frame Segment Listings	1-5
1.3.3 Selecting a Sample of Buildings from the Area Frame Segment Lists	1-6
1.4 LIST FRAME SAMPLING PROCEDURES	1-7
1.4.1 Obtaining the List Frame Lists	1-9
1.4.2 Restricting the List Frame Lists to Geographic Areas	1-9
1.4.3 First-Stage Selection and Processing Procedures: Selecting Units for Screening	1-9
1.4.4 Second-Stage Selection and Processing Procedures: Selecting Units for Interviewing	1-10
1.5 MATERIALS PREPARATION	1-10
CHAPTER 2: PROCEDURES FOR AREA FRAME SAMPLE ...	2-1
2.1 SELECTING A SAMPLE OF AREA FRAME SEGMENTS	2-1
2.1.1 Subsampling	2-1
2.1.1.1 Grouping the Segments	2-2
2.1.1.2 Identifying Special Blocks	2-2
2.1.1.3 Subsampling Segments with Clusters	2-3
2.1.1.4 Splitting the Segments into Chunks	2-3
2.1.1.5 Evaluating the Chunking Scheme	2-4
2.1.1.6 Updating the Database	2-4

2.2.	DEVELOPING COMPUTERIZED LISTINGS OF COMMERCIAL BUILDINGS IN THE AREA FRAME SEGMENTS	2-4
2.2.1	Definitions of a Qualified Building for CBECS	2-5
	2.2.1.1 Target Population for the 1995 CBECS	2-5
	2.2.1.2 Qualified Buildings for Area Frame Segment Listings	2-6
2.2.2	Creating the CBECS Database	2-6
2.2.3	Transferring Listing Data from Paper to the Database	2-9
	2.2.3.1 Editing and Coding	2-10
	2.2.3.2 Keying the Data	2-10
	2.2.3.3 Adding Keyed Data to the Database	2-10
	2.2.3.4 Proofreading and Corrections	2-11
2.2.4	Preparing Maps	2-13
	2.2.4.1 Reblocking Procedures	2-13
2.2.5	Preparing Listing Update Forms	2-14
2.2.6	Recruiting and Training Interviewers	2-14
2.2.7	Sendout and Updates	2-15
2.2.8	Processing Materials Returned from the Field	2-16
	2.2.8.1 Initial Review	2-16
	2.2.8.2 Coding	2-16
	2.2.8.3 Data Entry, Proofreading, and Corrections	2-17
	2.2.8.4 Consistency Checks and Facility Review	2-20
2.3	SAMPLE SELECTION	2-21
2.3.1	Eliminating Area Frame Cases That Overlap with the List Frame	2-22
2.3.2	Setting Final Sampling Rates	2-22
2.3.3	Selecting the Area Frame Sample	2-23
2.3.4	Post Sample Elimination of Overlap of Area Frame with List Frame	2-25

2.4	PREPARING MATERIALS FOR DATA COLLECTION	2-26
2.4.1	Preparing Listing Update Forms	2-26
2.4.2	Preparing Sketches and Maps	2-26
2.4.3	Creating the Selected Buildings File	2-27

CHAPTER 3: LIST FRAME SAMPLE PROCEDURES

3.1	DEVELOPING LIST FRAME LISTINGS	3-1
3.1.1	Government Facilities	3-2
3.1.2	Colleges and Universities	3-2
3.1.3	Hospitals	3-3
3.1.4	Pre-1986 Other Large Building List	3-3
3.1.5	Post-1986 Dodge Major Projects File	3-3
3.1.6	List Frame Integration	3-4
3.2	SELECTING A SAMPLE OF ESTABLISHMENTS FROM THE LIST FRAME LISTINGS	3-4
3.3	SCREENING SELECTED ESTABLISHMENTS	3-10
3.3.1	Removing Duplicate Records	3-10
3.3.2	Preparing Screening Materials	3-10
3.3.3	Conducting the Screening Phase	3-11
3.3.4	Processing Screened Facilities	3-12
	3.3.4.1 Coding the Screening Status	3-12
	3.3.4.2 Coding Building Records	3-13
3.3.5	Updating the Database	3-13
3.4	LIST FRAME BUILDING-LEVEL SAMPLE SELECTION	3-14
3.5	PREPARING LISTING UPDATE FORMS AND THE SELECTED BUILDINGS FILE	3-17

LIST OF TABLES

Table 2-1:	Final Area Frame Sampling Rates	2-23
Table 2-2:	Summary of Area Frame Sample Characteristics	2-24
Table 3-1:	List Frame Source Files Records Eligible for Facility Level Selection by Source and Estimated Square Footage Within 129 PSUs	3-3
Table 3-2:	1995 Overall Desired Sampling Rates for Facility Size Class 6, 7, 8, and 9	3-4
Table 3-3:	Total Facilities Selected From Each List Frame Source Files by Source and Estimated Square Footage Within 129 PSUs	3-5
Table 3-4:	Summary of the List Frame Processing Steps and Sample Selection	3-15

CHAPTER 1

INTRODUCTION AND OVERVIEW

CHAPTER 1: INTRODUCTION AND OVERVIEW

There have been important changes in the CBECS sample design between the 1992 CBECS and the 1995 CBECS. For 1995, both the area and list frames were updated. Unlike the 1989 and 1992 surveys in which partial area frame updates were conducted, the area frame listings were completely updated for 1995. In addition, new lists were obtained for 1995 to revise the list frame, including new lists of: federal facilities, hospitals, and colleges and universities.

For 1995 there were some important changes in the sample frame development and sampling procedures. These changes included:

- 1) Modifying the population for the study to exclude certain buildings and buildings on certain facilities
- 2) Restricting the scope of area frame listings from nonresidential buildings to commercial buildings
- 3) Reducing the number and scope of list frame lists to more narrowly focus on large buildings (i.e., buildings over 250,000 square feet)
- 4) Changing the frame development procedures to reduce the overlap between the area frame and the list frame
- 5) Modifying the selection probabilities to increase the take of large and moderate sized-buildings in the sample

The changes implemented for the 1995 CBECS were targeted to increase the efficiency of the survey and should reduce costs for future surveys.

1.1 CBECS SAMPLE DESIGN CONSIDERATIONS

The object of the CBECS sample frame development and sample selection procedures is to select a sample of buildings that adequately represents the population of commercial buildings in the United States (i.e., 50 states and the District of Columbia). Three characteristics of the commercial building sector have a significant impact on the sample design.

- 1) While there are many lists of commercial *establishments*, there is no comprehensive list of commercial *buildings*.
- 2) A small number of large commercial buildings consume a large share of the total energy used in the commercial building sector.
- 3) Many large buildings in the commercial building sector are part of multibuilding institutions and/or multibuilding facilities.

Since there is no consistent and comprehensive list of buildings throughout the United States, area frame sampling procedures were used to develop the core CBECS sample frame. In 1986, a multistage procedure was used to select 509 area frame segments for field listing. For the 1989 and 1992 surveys, 486 of the 509 area frame segments were used; a subsample of 191 segments (referred to henceforth as update segments) were updated to represent new construction. In 1995,

some of the segments were modified and the segment listings for all 509 segments were updated. The area frame segment listings furnish a representative sample of commercial buildings. The final CBECS area frame sample is a subsample of the buildings listed in the area frame segments.

Since large and energy intensive buildings consume a large share of the energy used in the commercial buildings sector, variance analyses show that the variances of energy consumption estimates are minimized by sampling large and energy intensive buildings at higher rates than small building and buildings that use energy less intensively. There are three ways that the sample frame development and sampling procedures accommodated the need to sample larger buildings at higher rates.

- 1) In the initial development of area frame segments, the U.S. conceptually was divided into segments that were equal in terms of *energy usage in commercial buildings*, rather than in terms of *number of buildings*. This adjustment served to reduce the variance of energy consumption at the segment level.
- 2) In the final area frame sample, larger and more energy intensive buildings were sampled at higher rates.
- 3) In addition to the area frame sample, a supplemental list frame of large buildings (i.e., buildings over 250,000 square feet) was developed.

In 1986, the CBECS supplemental list frame was developed using both existing building lists and a newly developed list of large buildings. The pre-1986 large building list was developed using a variety of sources, including publications listing certain building types (such as hotels, malls, and other public buildings) and by contacting government officials and other informed sources in the target areas. In addition, the 1986 list frame development procedures used national lists of government facilities, post offices, hospitals, colleges and universities, and noncollege schools.

Since many large buildings are part of multibuilding institutions and or facilities, a multistage procedure was required to develop a final list frame sample of large buildings. For example, the federal government list furnished a listing of all federal government facilities; the colleges and universities list furnished a listing of all colleges and universities. Lists of buildings could be obtained only through direct contact with the manager of the institution or the facility. Building lists were developed by first selecting a sample of institutions and/or facilities using an estimate of the square footage of buildings in the institution, and then obtaining a listing of buildings through telephone or in-field screening procedures.

For 1995, the list frame procedures were modified substantially. The pre-1986 large building list was reused. However, new lists were obtained for the other commercial buildings sector, including updates of the federal government facilities list, the hospitals list, and the list of colleges and universities. The noncollege schools list was dropped for 1995. Lists of construction projects were obtained from F.W. Dodge to furnish a list of other large buildings built after 1985.

1.2 DESCRIPTION OF THE UNIVERSE

The target population for the 1995 CBECS consisted of all commercial buildings larger than 1,000 square feet. To be eligible for the survey, a building had to satisfy three criteria: (1) it had to meet the definition of a building, (2) it had to be used primarily for some commercial purpose, and (3) it had to measure 1,001 square feet or more.

The first eligibility criterion, the building definition, was modified slightly for the 1995 CBECS. A building is defined as a structure totally enclosed by walls that extend from the foundation to the roof that is intended for human access. Structures such as water, radio, and television towers were excluded from the survey. Also excluded were partially open structures, such as lumber storage buildings; enclosed structures that people usually do not enter, such as pumping stations and cooling towers at electric power plants; enclosed structures that are not buildings, such as oil tanks, statues, and monuments; and dilapidated or uncompleted buildings missing a roof or a wall. In previous CBECS studies, parking garages were included in the study, despite not being "totally enclosed by walls." For 1995, buildings where more than 50% of the space was used for a parking garage were excluded, even if they were "totally enclosed by walls." Structures on pillars continued to be included in the definition of a commercial building.

The second criterion -- used primarily for some commercial purpose -- has been tightened in the successive CBECS surveys, and was further restricted for the 1995 CBECS. The initial target for the 1986 NBECS was *nonresidential* buildings. However, in the final implementation of the 1986 NBECS and of the CBECS surveys administered since that time, the target has been *commercial* buildings. For the 1995 CBECS, the target population was restricted to exclude commercial buildings on manufacturing establishments (e.g., office buildings). In past CBECS surveys, these buildings were included because they were commercial in nature. However, since these buildings are covered in the MECS survey, they were eliminated from the 1995 CBECS. Commercial buildings on other industrial (nonmanufacturing) establishments were still included in the CBECS (e.g., the office building at a sewage treatment facility).

Nonresidential buildings not included in the 1995 CBECS include:

- Buildings on military bases
- Buildings where access is restricted for national security reasons
- Farm buildings
- Buildings where more than 50% of the space is used for residential purposes, including apartment buildings
- Buildings where more than 50% of the space is used for manufacturing or other industrial purposes
- Mobile homes that are not placed on a permanent foundation, even if the mobile home is used for a commercial purpose

The third criterion -- size -- was added after fielding the 1986 NBECS to eliminate a large, inherently ill-defined group of marginal buildings which represent a very small fraction of the total commercial floorspace and energy consumption.

1.3 AREA FRAME SAMPLING PROCEDURES

The purpose of the area frame sampling procedures was to develop a sample of commercial buildings that are representative of all commercial buildings in the U.S. This was achieved through a three-step process, including:

- 1) Selecting a sample of geographic areas -- called area frame segments
- 2) Developing listings of all qualifying buildings in the area frame segments
- 3) Selecting a sample of buildings listed in the segments

The final sample of buildings is representative of eligible commercial buildings in the U.S..

1.3.1 Selecting Area Frame Segments

In 1986, multistage procedures were used to select the area frame segments. The target for these segments was to cover between 1/50th and 1/100th of the land area of the U.S.. The 509 area frame segments selected in 1986 were estimated to cover approximately 1/85th of the land area of the U.S.

- 1) The first stage of the sampling procedures, selection of Primary Sampling Units, was implemented simultaneously with the selection of the PSUs for the Residential Energy Consumption Survey (RECS) in 1983. With a few exceptions, the CBECS PSUs were identical to the RECS PSUs.

The approximately 3,100 counties and independent cities of the U.S. were grouped into 1,799 PSUs, consisting of one or more contiguous counties. PSUs with similar characteristics were grouped into 129 strata. (No PSU represented less than 2% of the stratum total.) One PSU (county or group of counties) was selected from each sample strata.

Thirty-two PSUs had populations large enough that each of these PSUs formed a stratum by itself. Those PSUs were considered to be self-representing (i.e., had a probability of selection of one). The probability of selection for the remaining PSUs was computed as the estimated size of the PSU divided by the estimated size of the stratum.

- 2) In the second stage of the sampling procedures, a sample of intermediate ZIP code groups was selected. It was estimated that there were 4,326,500 commercial buildings in the U.S. in 1984. In the second stage, we defined 43,265 conceptual segments of 100 "average-sized" buildings, estimated the number of segments included in each ZIP code group, and then selected the ZIP groups from which the final sample of segments would be selected.

The 1979 NBECS was used to develop models to relate commercial employment to square footage. These models were normalized to establish a correspondence between the 1983 estimated employment and the 1984 estimated number of buildings. The 1983 CBP data were used to estimate a measure-of-size (i.e., number of buildings) for individual ZIP code areas. ZIP code areas were grouped

into about 3,900 ZIP groups; each ZIP group contained at least one conceptual segment. A sample of 509 segments were selected within the ZIP group areas. Since some ZIP code areas had a large number of conceptual segments (i.e., had a lot of employment), the 509 segments were located within 444 ZIP group areas.

- 3) In the third stage of selection, individual segments were selected from the ZIP group areas using field counting procedures. A field interviewer established the final ZIP group boundaries and counted the number of buildings in each block of the selected ZIP group. Office staff estimated the total square footage of buildings in the ZIP group and divided the ZIP group into the number of conceptual segments represented by ZIP group according the CBP data. Finally, office staff randomly selected the required number of segments from each counted ZIP group area. (Note: For about 30% of the segments, the estimated number of conceptual segments was three or fewer. In these areas, the segment selection procedure was skipped and the entire ZIP group was treated as the segment.)

In 1986, all 509 segments were used in the sampling procedure. In 1989 and 1992, only 486 of the 509 were used. For 1995, all 509 segments were again used.

In 1989, the 191 update segments were reviewed to assess whether subsampling was appropriate. The primary targets for subsampling were segments where the area listed in 1986 covered two or more conceptual segments. Segments that contained large numbers of buildings also were considered for subsampling. In 1995, the remaining 313 segments were reviewed to assess whether subsampling was appropriate, using similar decision criteria to those employed in 1989.

1.3.2 Developing Area Frame Segment Listings

In 1986, the listing area was blocked (i.e., each block was assigned a reference number) and field listers were assigned to develop a complete listing of nonresidential buildings in the defined segments. Paper listings of buildings were developed for each block in each segment. The listing for each building furnished size/use class information for the building and building locating information.

In 1986, listers were instructed to list all nonresidential buildings, including many that were not "in-scope" in the final 1986 NBECS interviewing phase. The buildings excluded from the 1986 CBECS listings were:

- 1) Buildings that were completely residential (note that group quarters buildings are considered nonresidential)
- 2) Single family homes, even if they contained an office or business
- 3) Farms and farm buildings
- 4) Buildings located on military bases or reservations
- 5) Buildings where access is restricted for national security purposes

The 1986 CBECS listings contained a large number of ineligible buildings. During the CBECS interview, buildings where more than 50% of the floorspace was used for some noncommercial purpose (i.e., industrial, residential, or agricultural) were eliminated as ineligible. In addition, after the 1986 CBECS interviews were completed, buildings smaller than 1,000 square feet were eliminated for reporting purposes.

In 1989 and 1992, 196 of the 486 segments used for those surveys received listing updates. In these updates, buildings under 500 square feet were eliminated from the listings.

In 1995, all 509 area frame segments were updated. In addition, the area frame segment listings were restricted to eliminate a larger share of ineligible buildings. Buildings deleted from or not listed in the 1995 listings included:

- 1) Buildings where more than 75% of the floorspace was used for residential, agricultural, or industrial purposes.
- 2) Buildings that were less than 500 square feet.
- 3) Parking garages
- 4) Buildings on manufacturing establishments

We estimate that the area frame building ineligibility rate fell from about 30% in 1992 to about 15% in 1995.

1.3.3 Selecting a Sample of Buildings from the Area Frame Segment Lists

The target selection probabilities were equal for all buildings in each size/use strata. For 1995, there were 18 size/use strata used in the area frame. Larger buildings and buildings that used energy more intensely were sampled at higher rates. Actual selection probabilities for individual buildings were lower than the targets when the probability of listing a building was lower than the target probability of selection for a building that size/use class. (See Section 2 for a listing of the area frame selection probabilities).

For the 1995 CBECS sample, an attempt was made to eliminate the overlap between the area frame and the list frame prior to fielding the survey. Overlap analysis was conducted in two phases. First, target area frame buildings and facilities were examined to assess whether they appeared in the list frame. Second, a broader group of sampled area frame buildings were examined for possible overlap with the list frame. Overlapping buildings were not eliminated from the area frame listings. Rather, they received a code that indicated that, if they were selected for interview, they should not be fielded.

1.4 LIST FRAME SAMPLING PROCEDURES

The list frame was first developed for the 1986 NBECS. For 1995, the list frame was updated and revised in a number of ways.

The list frame can be used for several different, but complementary, purposes.

- 1) **Increase Sample of Large Buildings:** Since the area frame segment listings cover about 1/85th of the geographic area in the country, the expected proportion of large buildings available to be selected in the area frame is about 1/85th. The target selection probabilities for buildings over 250,000 is greater than 1/85th. The list frame can be used to increase the total sample of large buildings.
- 2) **Increase Sample of Special Use Buildings:** EIA may wish to examine special kinds of buildings, such as government buildings, education buildings, and/or health care buildings. The list frame can be used to increase the total sample of special types of buildings.
- 3) **Reduce Clustering for Certain Types of Buildings:** Certain types of buildings are highly clustered. For example, colleges and universities are located on central campuses. In the area frame segments, we would expect to get about 1/85th of the colleges. In many area frame segments that include colleges, most of the segment is made up of the college. On the other hand, most noncollege segments include many different kinds of buildings. Using the area frame only, all college buildings are represented by buildings from relatively few colleges. Using the list frame approach, all college buildings can be represented by a large number of colleges.
- 4) **Reduce Variance in Estimates for Certain Types of Buildings:** Certain types of buildings were not well-represented in the original MOS. In particular, federal government employment is not covered by the CBP and, therefore, federal government buildings were not part of the estimated MOS for a ZIP code area. Using that MOS, the expected mean number of federal government buildings estimated is still unbiased. However, the variability in the estimated number of government buildings will be larger than for other building types, particularly if government buildings are more clustered than other building types. Using a list frame with government buildings will reduce the variability of the estimated number of large buildings.

In 1986, the list frame was developed using both a large buildings list and special lists of building types. The special lists served two purposes. First, they reduced the amount of work required to develop the large building list, since certain kinds of buildings could be ignored in the large building list work. Second, the special lists offered the opportunity to sample certain types of buildings at a higher rate.

In 1986, the large building list was targeted to include all buildings over 250,000 square feet in MSA PSUs and over 100,000 square feet in nonMSA PSUs. The special lists were targeted to include all establishments where the total square footage of all buildings was over 50,000 square feet.

In 1986, the actual list frame selection procedures achieved goals 1, 3, and 4; the number of large buildings was increased; the sample of clustered building types was "disperse," and the variance in estimates for government buildings was reduced. Goal 2 was not achieved because special building types were not sampled at higher rates than other buildings.

For 1995, we reviewed the goals including a supplemental list frame. We found that the same goals (1, 3, and 4) could be achieved with a different list frame structure. We implemented the following changes.

- 1) Large Building List Targets: The large building list frame was limited to buildings over 250,000 square feet in all PSUs. Optimal allocation analyses suggested that there was no significant benefit in using a lower size cutoff for nonMSA PSUs.
- 2) Elimination of the Noncollege Schools List: Analysis of 1986 data showed that 68% of the facilities selected from this list were smaller than 250,000 square feet. Since the expectation is that this group of buildings is well-covered in the area frame, the list frame was duplicating, rather than supplementing the area frame in this area.
- 3) Other Special List Targets: Minimum sizes were raised for the remaining special lists, since selection of small facilities was expected to duplicate area frame coverage. The minimum size of facilities from the hospital list was 250,000 square feet. Minimum size of facilities from the Colleges and University list was 1,000,000 square feet.
- 4) Special Government Building Targets: New data available on the government facility list allowed the government list to be further limited. The data show the number of buildings and gross square footage of buildings on government facilities by building type. The special government list was limited to facilities that had any building types with more than 250,000 square feet of floorspace.

In addition, we changed the list frame "take all" rule. In previous surveys, for any facility with three or fewer buildings, all buildings on the facility were considered to be part of the sample. As a result, a number of small buildings on multibuilding facilities were sampled at rates that were substantially higher than the target selection rates for buildings in their size/use class. For the 1995 survey, all list frame buildings were subsampled after screening.

The final list frame, then, was comprised of the following.

- 1) Government Buildings: A list of government facilities with at least one building type that had at least 250,000 square feet of floorspace.
- 2) Colleges and Universities: A list of colleges and universities estimated to have at least 1,000,000 square feet of floorspace.
- 3) Hospitals: A list of hospitals estimated to have over 250,000 square feet of floorspace.

- 4) Large Buildings: A list of buildings over 250,000 square feet developed in 1986.
- 5) Dodge Projects: A list of Dodge projects over 250,000 square feet for the years 1986 through 1994.

1.4.1 Obtaining the List Frame Lists

The list frame lists were obtained from a variety of sources described in Section 3. For each list, we obtained the most "up-to-date" version that was available in early 1995.

1.4.2 Restricting the List Frame Lists to Geographic Areas

The list frames were restricted to the PSUs selected for the area frame. List frame selections were then located in the same general areas as the area frame selections. Generally, interviewers could schedule interviews at list frame buildings in a way that integrated the area frame and list frame work.

1.4.3 First-Stage Selection and Processing Procedures: Selecting Units for Screening

Using the restricted list frames, a sample of units was selected for screening. The selection probabilities are listed in Section 3. In most cases, the selected unit represented a facility (i.e., a group of related buildings). The following procedures were applied to each unit.

- 1) Overlap Analysis: The list frame conceptually was integrated into a single frame by eliminating the overlap between list frame lists. Rather than conduct a comprehensive list-to-list match to eliminate all overlap, a sample-to-list procedure was used. In this procedure a list hierarchy was developed. A unit selected from a list was checked for inclusion in all higher ranking lists. If it was found on a higher ranking list, it was deleted from the sample.
- 2) Screening: The list frame unit was contacted and screened to establish unit eligibility and to obtain a list of buildings. For each list frame unit, we first assessed whether the unit was out-of-scope because it was a manufacturing facility. If the unit was in-scope, we then obtained a list of all eligible buildings (i.e., we eliminated buildings under 1,000 square feet, parking garages, industrial buildings, residential buildings, and agricultural buildings). Finally, we obtained information that would facilitate the sample-to-frame overlap analysis. (For example, when screening a hospital, we asked whether it was owned by the federal government and whether it was part of a college or university campus. If there was a positive response to either question, we checked the government frame or the college/university frame once again for overlap.)

Screened lists of buildings for each unit were put into a database. Facilities where screening could not be completed were sent to Westat for field screening.

1.4.4 Second-Stage Selection and Processing Procedures: Selecting Units for Interviewing

Each building on the screened lists was sampled using the building selection probabilities outlined in Section 3. In previous CBECS studies, if there were three or fewer buildings associated with a facility, all buildings in the facility were interviewed. For the 1995 CBECS, buildings in all facilities were subsampled after screening using the building-level selection probabilities.

1.5 MATERIALS PREPARATION

For the area frame, the following materials were delivered to Westat.

- 1) Selected buildings file -- containing locating and sampling information
- 2) Segment maps
- 3) Block listings of buildings for all blocks in which a building was selected for interview
- 4) Sketches for the entire segment

For the list frame, the following materials were delivered to Westat.

- 1) Selected buildings file -- containing locating and sampling information
- 2) Facility listing form -- with information on the facility, eligible buildings in the facility, and selected buildings in the facility
- 3) Facility-level telephone screening forms

CHAPTER 2

PROCEDURES FOR AREA FRAME SAM- PLE

CHAPTER 2: PROCEDURES FOR AREA FRAME SAMPLE

The area frame sampling procedures are designed to select a sample of qualifying commercial buildings in the 50 states and the District of Columbia. The steps in selecting the area frame sample are:

- 1) Selecting a sample of area frame segments
- 2) Developing computerized listings of qualified commercial buildings in the segments
- 3) Selecting a sample of buildings from the segment listings
- 4) Preparing CBECS interviewer locating materials

A detailed description of each step is provided in this section. In addition, relevant reports, instructions, and materials are included in the appendices.

2.1 SELECTING A SAMPLE OF AREA FRAME SEGMENTS

A multistage procedure was used to select the sample of area frame segments. The selection stages included:

- 1) Selection of Primary Sampling Units (PSUs)
- 2) Selection of ZIP Code Groups
- 3) Selection of Secondary Sampling Units (SSUs) or Segments
- 4) Segment Subsampling

CBECS PSUs were selected in 1983 as part of the RECS/CBECS sample development procedures. These procedures are documented in the **1984 RECS Sample Design Procedures Manual** and the **1986 NBECS Sample Selection Procedures Manual**.

CBECS ZIP Code groups and SSUs (segments) were selected in 1986 as part of the 1986 NBECS sample development procedures. These procedures are documented in the **1986 NBECS Sample Selection Procedures Manual**.

Segment subsampling for the segments that were updated in 1989 are documented in the **1989 NBECS Sample Selection Procedures Manual**. Segment subsampling for the remaining segments are documented here.

2.1.1 Subsampling

The primary reason from subsampling segments was that about 35 percent of the segments were larger than the targeted segment size. In 1986 these segments were left at the larger sizes to reduce listing costs. However, the segment listing can be subsampled, thereby reducing relisting costs.

The guiding principle of the subsampling procedure was to create a subsampled segment that has a similar structure to the initial segment in terms of the distribution of buildings by size. While other subsampling procedures were acceptable, this conservative approach is expected to minimize the differences between survey frames.

The subsampling procedure broke down large segments into smaller parts, called chunks, which are subsets of the blocks that make up individual segments. Each segment was evaluated to determine if it was a good candidate for subsampling. Segments that were deemed to be good candidates were divided into chunks, and the chunks were evaluated to find out if the subsampling procedure was successful. The selected chunk represented the subsampled portion of the segment. Segments that were not good candidates for subsampling and segments where subsampling was not successful were left intact. Roughly 30% of the 318 nonupdate segments were subsampled. The sections below describe the general procedures involved in the subsampling process. The Subsampling Procedures document (Item B-1 in the appendix) provides complete specifications for this process. The Results of Subsampling document (Item B-3) contains the results, by group, of the subsampling effort.

2.1.1.1 Grouping the Segments

The 318 segments were divided into eight mutually exclusive groups (1A, 1B, 2A, 2B, 3A, 3B, 4A, 4B), according to their MOS (measure of size), number of blocks, and number of buildings. Each group was evaluated to determine whether or not the segments should be subsampled and, if so, what specific method should be used.

- Purpose: The purpose of this program is to assign each segment to one of eight groups. The program also calculates segment and block-level summary variables needed for the chunking process.
- Input(s): CN6616.PRO.DRAFT.CBECS92.AREA.FINAL.NOTUP
CN6616.PRO.DRAFT.CBECS95.SUBSAMP.BATCH86
- Output(s): CN6616.PRO.DRAFT.CBECS95.SUBSAMP.SEGS:SEGLEV
CN6616.PRO.DRAFT.CBECS95.SUBSAMP.SEGS:GROUP1A
CN6616.PRO.DRAFT.CBECS95.SUBSAMP.SEGS:GROUP1B
CN6616.PRO.DRAFT.CBECS95.SUBSAMP.SEGS:GROUP2A
CN6616.PRO.DRAFT.CBECS95.SUBSAMP.SEGS:GROUP2B
CN6616.PRO.DRAFT.CBECS95.SUBSAMP.SEGS:GROUP3A
CN6616.PRO.DRAFT.CBECS95.SUBSAMP.SEGS:GROUP3B
CN6616.PRO.DRAFT.CBECS95.SUBSAMP.SEGS:GROUP4A
CN6616.PRO.DRAFT.CBECS95.SUBSAMP.SEGS:GROUP4B
- Program: CN6616.PRO.CBECS95.SUBSAMP:GROUPS

2.1.1.2 Identifying Special Blocks

Special blocks are those blocks that contain a high percentage of buildings or large buildings compared to the rest of the blocks in the segment. The special blocks were excluded from the subsampling routine, but were added to the selected chunk as part of the subsampled area. The program SPECBLKS identified special blocks in groups 1A and 2A. The programs GP3ASPEC

AND GP4ASPEC identified special blocks in groups 3A and 4A, respectively. The following is a generic program description for all three programs.

- Procedure: This program flags each special block in each segment.
- Input(s): CN6616.PRO.DRAFT.CBECS95.SUBSAMP.SEGS:GROUPXX
(where XX is the group number)
- Output(s): CN6616.PRO.DRAFT.CBECS95.SUBSAMP.SEGS:GROUPXX
(where XX is the group number)
- Program: CN6616.PRO.CBECS95.SUBSAMP:XXXXXXXXXX
(where XXXXXXXXX is the program name)

2.1.1.3 Subsampling Segments with Clusters

Segments with MOS>2 were examined to see if they contained a cluster or clusters. A cluster is a group of 10 or fewer contiguous blocks that contain at least 50% of the total buildings, total weighted buildings, or total buildings of at least 25,000 sq. ft. The program SETCLUST was written to flag blocks that were part of a cluster. These blocks were chunked separately from the non-clustered blocks in that segment; the subsampled portion of the segment consisted of a selected chunk of the non-clustered blocks and a selected chunk of clustered blocks.

- Procedure: This program will flag each block that is part of a cluster.
- Input(s): CN6616.PRO.DRAFT.CBECS95.SUBSAMP.SEGS:GROUPxx
(where xx is the group number)
- Output(s): CN6616.PRO.DRAFT.CBECS95.SUBSAMP.SEGS:GROUPxx
(where xx is the group number)
- Program: CN6616.PRO.CBECS95.SUBSAMP:SETCLUST

2.1.1.4 Splitting the Segments into Chunks

Each subsampled segment was split into chunks, the number of chunks being equal to the MOS for that segment. The chunks consisted of numerically sequential blocks balanced with respect to building count, weighted building count, and the count of buildings at least 25,000 sq. ft. Each GROUPXX program run performed a pre-specified number of iterations of the chunking routine for the specified segment. The program was modified before each run to pass the segment number, number of blocks in the segment, and number of chunking schemes needed. Each iteration of the chunking routine randomly selected one of the blocks as a starting point, thereby generating a number of different chunking schemes. In general, 40 iterations of the chunking routine were specified for each program run, thereby ensuring a large enough pool of chunking schemes from which to choose. Refer to Item B-2 in the appendix for a sample copy of a chunking program printout.

- Procedure: This program split each segment into an assigned number of chunks. The program run produced up to 40 chunking schemes for each segment. The program also produces summary statistics for each chunking scheme. Each group chunking program was tailored to take into account the unique properties of that group, such as MOS and clustering.

Input(s): CN6616.PRO.DRAFT.CBECS95.SUBSAMP.SEGS:GROUPXX
(where XX is the group number for the selected segment)
User input

Output(s): A printed listing of the forty chunking schemes and summary statistics for each scheme.

Program: CN6616.PRO.CBECS95.SUBSAMP:GROUPxx
(where xx is the group number for the selected segment)

2.1.1.5 Evaluating the Chunking Scheme

Each chunking scheme for a particular segment was evaluated to see which scheme yielded the best results. A chunk (or, in some instances, chunks) was selected from this scheme, and this chunk became the subsampled portion of the segment.

2.1.1.6 Updating the Database

The results of the subsampling procedure were recorded on the Segment Subsampling Results form (Item B-4 in the appendix) and placed in the subsampling folder. The results were also recorded on Results of Subsampling document. The Results of Subsampling document was used to manually update the database. Further processing and updating was done only to that selected portion of the segment.

2.2 DEVELOPING COMPUTERIZED LISTINGS OF COMMERCIAL BUILDINGS IN THE AREA FRAME SEGMENTS

The goal of the 1995 procedures was to develop an up-to-date listing of qualified commercial buildings in the 509 area frame segments. Since these segments had been used for the 1986, 1989, and 1992 surveys, the listings were partially complete in all segments. In addition, since some segments (191) had been updated in 1989 and 1992, while others had not been updated since 1986 (318), the materials for different segments were in different formats. For 1995, a series of procedures were employed to develop a single database with listings for all segments. The procedures included:

- Developing the CBECS Address Processing System (CAPS) database for all segments
- Coding, keying, and proofreading the 1986 listings for the 318 non-update segments
- Reviewing and updating orientation and segment maps for all 509 segments

- Preparing maps, special sketches, listing update booklets, and listing materials for all segments
- Developing training materials and training field listers
- Updating listings, including: recording changes and/or additions to the area frame listings, relisting segments or parts of segments when required, and updating orientation and segment maps
- Coding, keying, and proofreading updated information
- Performing consistency checks
- Selecting sample
- Preparing materials for data collection

2.2.1 Definition of a Qualified Building for CBECS

When the CBECS sample frame was developed in 1986, the goal was to develop a sample frame for nonresidential buildings. Over time, CBECS has become more narrowly focused on commercial buildings. For 1995, the area frame segment listing procedures attempted to restrict listings to qualified commercial buildings.

2.2.1.1 Target Population for the 1995 CBECS

In general, a building is eligible for interview in CBECS if it meets three criteria: (1) it meets the survey's definition of a building, (2) it is used primarily for some commercial purpose, and (3) it measures more than 1,000 square feet. For the 1995 survey, the following additional rules are designed to implement these general guidelines.

- 1) With respect to criteria (1), parking garages sometimes do not meet the definition of a building (i.e., often they are not enclosed structures). Because these structures use relatively little energy and do not "fit" the CBECS building interview in many cases, parking garages have been deleted from the target population for the 1995 sample, even if they fit the definition of a building. Any structure where more than 50% of the floor space is used as a parking garage is excluded from the CBECS.
- 2) With respect to criteria (2), commercial buildings that are part of a facility occupied by a manufacturing establishment have been deleted to eliminate the overlap between CBECS and MECS. The MECS sample is selected from a sample frame of manufacturing establishments. The MECS data collection obtains information regarding the amount of energy used by all buildings and other structures associated with the establishment, including those whose primary function might be considered to be commercial (e.g., office space).

- 3) With respect to criteria (2), a building is considered to be used primarily for some commercial purpose if more than 50% of its floor space is used for a commercial purpose. Thus, buildings where 50% or more of the floor space is used for an agricultural, residential, or industrial use are excluded.¹

In addition, the target population for the 1995 CBECS excludes buildings on military bases and other buildings where access is restricted for national security reasons.

The target population for the 1995 CBECS differs in two ways from the target population for the 1986, 1989, and 1992 studies. First, parking garages were included in the previous studies. Second, commercial buildings on manufacturing establishments were included in previous studies. These changes should simplify the fielding procedures. In particular, commercial buildings that are part of a manufacturing establishment present two problems. First, many manufacturing establishments are located on secure facilities and the CBECS interviewer must gain special permission to enter the facility. Second, many manufacturing establishments have a single meter for the entire facility, thereby requiring CBECS analysts to use "allocation" procedures to estimate energy usage. The impact of these changes on CBECS estimates should be modest, since both types of facilities represent a small part of the total population of commercial buildings.

2.2.1.2 Qualified Buildings for Area Frame Segment Listings

In general, the goal of the field listings procedures was to list only qualified commercial buildings. However, in some cases it is difficult to determine the primary use of a building (i.e., the use that occupies more than 50% of the floor space.) In addition, it is difficult to assess the precise measurement of a building. For these reasons, the listing procedures asked field listers to use the following, less restrictive guidelines.

- 1) The minimum size for listing was reduced from 1,000 square feet to 500 square feet
- 2) Buildings where at least 25% of the floor space was used for commercial purposes were included in the listings

While these guidelines increased the number of ineligible buildings in the sample, they helped to ensure that all eligible buildings would be listed.

2.2.2 Creating the CBECS Database

The CBECS database held all data about each segment, including segment-level, block-level, facility-level, and building-level information. The database also contained historical information about each level, such as changes in line numbers and block numbers over the course of the CBECS studies. The basis of this database was the mainframe files that had been maintained over the previous studies.

¹ If space in a farm or agricultural building is used for retail sales to the general public, the building is defined as commercial, rather than agricultural. Dormitories and other group quarters are considered to be commercial buildings.

The CBECS Area Frame Database Application is a Clipper-compiled executable program that is menu-driven. Users may edit or add information, print reports, and create files and field materials using the application. The program is located on a local area network (LAN) and is designed to have multiple concurrent users. The application locks files and records when necessary to prevent a system failure from attempted simultaneous updates.

The database includes the segment, block, building, and facility tables. The information for these tables came from multiple sources, but the primary files were the 1992 CBECS Area Frame file for non-update segments and the 1992 CBECS Area Frame file for update segments. These files were downloaded from their storage site on the DOE mainframe computer, and the data were translated by SAS programs into a format suitable for database tables. Further block, building, and facility address information was retrieved from the PC-based NAPS (NBECS Address Processing System) files for the updated segments. Detailed memos outlining this process can be found in the appendix, Items C-1 - C-4. The CBECS database table structure and a description of the menu options and called functions (Items C-5 and C-6, respectively) can also be found in the appendix.

- Procedure: This SAS program read in mainframe file for segments that were not updated in previous CBECS. It separated this hierarchical flat file into multiple datasets, one for each level (segment, block, and building). It extracted those segments that were updated and subsampled in 1989 and 1992 and put them aside in the NDUPES.TXT file, because they also exist in the mainframe updated segment file. If a building block or line number changed between 1986 and 1992, it created a historical record for that building. It then downloaded all the datasets into text files for later uploading to the database tables.

Input(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\NOTUP.TXT
(downloaded from DOE mainframe file CN6616.PRO.DRAFT.CBECS92.AREA.-FINAL.NOTUP)

Output(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\NSEGMENT.TXT
M:\START\PROJ\8800DOE\CBECS\DATABASE\NBUILDNG.TXT
M:\START\PROJ\8800DOE\CBECS\DATABASE\NBLOCK.TXT
M:\START\PROJ\8800DOE\CBECS\DATABASE\NHISTORY.TXT
M:\START\PROJ\8800DOE\CBECS\DATABASE\NDUPES.TXT

Program: M:\START\PROJ\8800DOE\CBECS\DATABASE\NOTUP.SAS

- Procedure: This SAS program read in the combined NAPS file with address information for all updated segments. It separated this hierarchical flat file into multiple datasets, one for each level (segment, block, building, and facility). For those blocks with special city or town information, it created records for the NAP_PO.TXT file that were processed later in a Clipper program. It read in the mainframe file for segments that were updated in previous CBECS. It separated this hierarchical flat file into multiple datasets, one for each level (segment, block, and building). The program also combined the segment level dataset from the mainframe file with the segment level dataset from NAPS, and created a text file from the result. It combined the block level dataset from the mainframe file with the block level dataset from NAPS, and created a text file from the result. It combined the

building level dataset from the mainframe file with the building level dataset from NAPS, and created a text file from the result. If a building block or line number changed between 1986, 1989, or 1992 CB ECS, it created historical records for that building. It downloaded all the datasets into text files for later uploading to the database tables.

Input(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\ALL.NPS
(all segment-level NAPS files combined into one file)
M:\START\PROJ\8800DOE\CBECS\DATABASE\UPDATE.TXT
(downloaded from DOE mainframe file CN6616.PRO.DRAFT.CBECS92.AREA.-FINAL.UPDATE)

Output(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\NAP_FACL.TXT
M:\START\PROJ\8800DOE\CBECS\DATABASE\NAP_PO.TXT
M:\START\PROJ\8800DOE\CBECS\DATABASE\USEGMENT.TXT
M:\START\PROJ\8800DOE\CBECS\DATABASE\UBUILDNG.TXT
M:\START\PROJ\8800DOE\CBECS\DATABASE\UBLOCK.TXT
M:\START\PROJ\8800DOE\CBECS\DATABASE\UHISTORY.TXT

Program: M:\START\PROJ\8800DOE\CBECS\DATABASE\NOTUP.SAS

- Purpose: This Clipper program took the text files created by NOTUP.SAS and appended them to the database tables.

Input: M:\START\PROJ\8800DOE\CBECS\DATABASE\NSEGMENT.TXT
M:\START\PROJ\8800DOE\CBECS\DATABASE\NBUILDNG.TXT
M:\START\PROJ\8800DOE\CBECS\DATABASE\NBLOCK.TXT
M:\START\PROJ\8800DOE\CBECS\DATABASE\NHISTORY.TXT

Output: M:\START\PROJ\8800DOE\CBECS\DATABASE\SEGMENT.DBF
M:\START\PROJ\8800DOE\CBECS\DATABASE\BUILDING.DBF
M:\START\PROJ\8800DOE\CBECS\DATABASE\BLOCK.DBF
M:\START\PROJ\8800DOE\CBECS\DATABASE\HISTORY.DBF

Program: M:\START\PROJ\8800DOE\CBECS\DATABASE\NOTUP.EXE

- Purpose: This Clipper program took the text files created by UPDATE.SAS and appended them to the database tables. Special block city and town information was written from NAP_PO.TXT to the BLOCK table.

Input: M:\START\PROJ\8800DOE\CBECS\DATABASE\USEGMENT.TXT
M:\START\PROJ\8800DOE\CBECS\DATABASE\UBUILDNG.TXT
M:\START\PROJ\8800DOE\CBECS\DATABASE\UBLOCK.TXT
M:\START\PROJ\8800DOE\CBECS\DATABASE\UHISTORY.TXT
M:\START\PROJ\8800DOE\CBECS\DATABASE\NAP_FACL.TXT
M:\START\PROJ\8800DOE\CBECS\DATABASE\NAP_PO.TXT

Output: M:\START\PROJ\8800DOE\CBECS\DATABASE\SEGMENT.DBF
M:\START\PROJ\8800DOE\CBECS\DATABASE\BUILDING.DBF
M:\START\PROJ\8800DOE\CBECS\DATABASE\BLOCK.DBF
M:\START\PROJ\8800DOE\CBECS\DATABASE\HISTORY.DBF
M:\START\PROJ\8800DOE\CBECS\DATABASE\FACILITY.DBF

Program: M:\START\PROJ\8800DOE\CBECS\DATABASE\UPDATE.EXE

- Purpose: This Clipper program set the T variables at both the segment and block level for all segments according to the memos in the appendix, Items C-1 - C-3.

Input: M:\START\PROJ\8800DOE\CBECS\DATABASE\SEGMENT.DBF
M:\START\PROJ\8800DOE\CBECS\DATABASE\BLOCK.DBF

Output: M:\START\PROJ\8800DOE\CBECS\DATABASE\SEGMENT.DBF (revised)
M:\START\PROJ\8800DOE\CBECS\DATABASE\BLOCK.DBF (revised)

Program: M:\START\PROJ\8800DOE\CBECS\DATABASE\SET_TVAR.EXE

- Purpose: This menu option allowed the NAPS information for selected segments to be written back out to text files in the hierarchical format. This was used for checking that the upload had been done correctly.

Input: M:\START\PROJ\8800DOE\CBECS\DATABASE\BUILDING.DBF
M:\START\PROJ\8800DOE\CBECS\DATABASE\BLOCK.DBF
M:\START\PROJ\8800DOE\CBECS\DATABASE\FACILITY.DBF

Output: M:\START\PROJ\8800DOE\CBECS\DATABASE\CHKNAPS\SXXXYY.TXT,
where XXX was the PSU and YY the field unit of the segment requested.

Program: M:\START\PROJ\8800DOE\CBECS\DATABASE\SET_TVAR.EXE

2.2.3 Transferring Listing Data from Paper to the Database

The best source of complete building and address information for the non-update segments was the original listings (on paper) completed in 1986. These listings were coded in preparation for keying, keyed into a flat file, and then added to the CBECS database. The information for each segment was printed for proofreading.

Before any of the preparation work for updating the area frame for the 1995 CBECS was initiated, all of the sampling materials from the 1986, 1989, and 1992 CBECS were retrieved from storage and moved to a central work area. The materials were consolidated into a single set of files for each SSU according to the specifications in the Preliminary Sample File Work (Item E-1 in the appendix); any sampling documents or materials that were superseded by similar materials were discarded.

2.2.3.1 Editing and Coding

The purpose of editing and coding was to prepare the listings for data entry. Building, facility, and Post Office/community name information that would serve as the initial input to the CBECS database and that would later be updated by fieldworkers was identified by coders in the coding/editing step. Coders prepared this information for data entry according to the instructions in "Preparing the 1986 Listings for Keying" (Item E-2 in the appendix.) For this task, the coders examined the handwritten building listings, looking for key pieces of information for each building. They highlighted, using different colored highlighters keyed to specific items of information, the building's name, address, and use, and any comments about the building that might aid an interviewer in identifying the building. Information was edited, as necessary, to read as the coder would like to see the information stored in the database. The "1995 CBECS CAPSWORK Building Record" (Item E-3) was used to rewrite building entries that were illegible on the listing form or that couldn't be clearly edited in the available margin space. Other information, such as segment number, block number, and line number, was also highlighted as detailed in the instructions. Coders also created facility entries when two or more buildings appeared to be "related". Facility information was coded onto the "1995 CBECS CAPSWORK Facility Summary Form" (Item E-4.) Post Office/community name records were coded when that information differed from the ZIP Code and Post Office data already in the database. The information was recorded on the "1995 CBECS CAPSWORK Post Office/Community Record" (Item E-5).

2.2.3.2 Keying the Data

All of the paperwork for a segment – the highlighted listing forms and completed coding forms – were clipped together. Batches of segments were delivered to the data entry contractor weekly. Information on the listing and coding forms was keyed according to the instructions in the memo dated November 30, 1994, "Specifications for Keying 1986 NBECS CAPSWORK File Records." A follow-up memo dated December 19, 1994, "Keying "Double-blue, Double-pink" Segments," addressed a special data entry situation involving ten segments that were highlighted before the final coding procedures were in place. Copies of these memos are found in the appendix as Item E-9 and Item E-10, respectively. The data entry contractor consulted the project manager on how to handle any ambiguous or incomplete work. Keyed data were 100% verified.

2.2.3.3 Adding Keyed Data to the Database

The keying service returned the completed CAPS files to Response Analysis on diskettes. A file usually contained several segments, and a segment was always completely keyed before being returned to RAC. Each file was uploaded to the database tables in two steps, a SAS program and a Clipper program.

In the first step of this procedure, the program CAPS.SAS split the hierarchical flat file furnished by the data entry subcontractor into several datasets corresponding to the database tables created for block-, facility-, and building-level records. The program also checked for duplicate records and errors in the facility record structure. Errors were checked and corrections made to the keyed file as necessary.

In step two, a Clipper program uploaded the three files to the database tables.

- Procedure: This SAS program read in the newly-keyed CAPS file. It divided the hierarchical flat file into multiple datasets, one for each level (block, building, and facility). It downloaded all the datasets into text files for later processing by a Clipper program.

Input(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\CAPS.TXT (keyed file sent from keying service)

Output(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\CAP_BLDG.TXT
M:\START\PROJ\8800DOE\CBECS\DATABASE\CAP_FACL.TXT
M:\START\PROJ\8800DOE\CBECS\DATABASE\CAP_PO.TXT

Program: M:\START\PROJ\8800DOE\CBECS\DATABASE\CAPS.SAS

- Procedure: This Clipper program uploaded the text files created by the previous program to the database tables. Address information was copied from the CAPS files to the already-existing building and facility records. Special block city and town information was copied, where found, to the block records from the CAP_PO.TXT file.

Input(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\CAP_BLDG.TXT
M:\START\PROJ\8800DOE\CBECS\DATABASE\CAP_FACL.TXT
M:\START\PROJ\8800DOE\CBECS\DATABASE\CAP_PO.TXT

Output(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\BUILDING.DBF
M:\START\PROJ\8800DOE\CBECS\DATABASE\BLOCK.DBF
M:\START\PROJ\8800DOE\CBECS\DATABASE\FACILITY.DBF

Program: M:\START\PROJ\8800DOE\CBECS\DATABASE\CAPS.EXE

2.2.3.4 Proofreading and Corrections

Information uploaded to the database was printed and compared to coded 1986 block listing forms. Building records, divided into their respective blocks, and facility records were printed for each segment. Editing staff checked for errors made during the keying process and scanned for mistakes made by coding staff during the highlighting phase. Corrections were made on the printed records in red ink and the building and facility records were updated in the database through an editing window.

- Purpose: This menu option prints out proofreading listings for buildings and facilities in those segments that have had the CAPS files uploaded.

Procedure: Select the menu option "Print CAPS for proofreading" under the "Data Cleaning" submenu under the "Listing" menu option of the CBECS Area Frame Database Application.

Input(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\STATUS.DBF
M:\START\PROJ\8800DOE\CBECS\DATABASE\BLOCK.DBF
M:\START\PROJ\8800DOE\CBECS\DATABASE\BUILDING.DBF

Output(s): Printed reports

- Purpose: This menu option allows the user to edit building information.

Procedure: Select the menu option "Buildings" under the "Data" menu option of the CBECS Area Frame Database Application.

Input(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\BUILDING.DBF
User input

Output(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\BUILDING.DBF (revised)

- Purpose: This menu option allows the user to edit block information.

Procedure: Select the menu option "Blocks" under the "Data" menu option of the CBECS Area Frame Database Application.

Input(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\BLOCK.DBF
User input

Output(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\BLOCK.DBF (revised)

- Purpose: This menu option allows the user to edit facility information.

Procedure: Select the menu option "Facilities" under the "Data" menu option of the CBECS Area Frame Database Application.

Input(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\FACILITY.DBF
User input

Output(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\FACILITY.DBF (revised)

- Purpose: This menu option allows the user to view facilities and their associated buildings on the same screen, and select any building or facility for editing.

Procedure: Select the menu option "Facilities and associated buildings" under the "Data" menu option of the CBECS Area Frame Database Application.

Input(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\BUILDING.DBF
M:\START\PROJ\8800DOE\CBECS\DATABASE\FACILITY.DBF
User input

Output(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\BUILDING.DBF (revised)
M:\START\PROJ\8800DOE\CBECS\DATABASE\FACILITY.DBF (revised)

- Purpose: This menu option allows the user to edit segment information.

Procedure: Select the menu option "Segments" under the "Data" menu option of the CBECS Area Frame Database Application.

Input(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\SEGMENT.DBF
User input

Output(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\SEGMENT.DBF (revised)

- Purpose: This menu option allows the user to mark that a selected segment has completed the proofreading and correcting process.

Procedure: Select the menu option "Enter status after proofing/correcting" under the "Data cleaning" submenu under the "Listing" menu option of the CBECS Area Frame Database Application.

Input(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\STATUS.DBF
User input

Output(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\STATUS.DBF (revised)

2.2.4 Preparing Maps

Orientation and segment maps were needed for each segment. The orientation map identified the general area in which the segment was found, and the segment map identified the individual blocks and streets within each segment. The maps in each segment folder were evaluated to see if they could be used again, or if new maps were needed. When new maps were needed, they were generated using the DeLorme MapExpert version 2.0 for Windows software according to the procedures discussed in the Instructions for CBECS mappers (Item D-1 in the appendix.)

2.2.4.1 Reblocking Procedures

During the course of mapping, it was sometimes necessary to split or combine existing blocks. This process, called reblocking, was carried out according to the specifications in the Instructions for CBECS mappers. The database was manually updated to reflect these changes.

2.2.5 Preparing Listing Update Forms

Listing update forms (LUFs) were generated by the CBECS database. Two copies were printed for each segment -- one to send to the fieldworker and the other to remain at Response Analysis for reference.

- Purpose: This menu option prints the Listing Update Forms for the selected segment

Procedure: Select the menu option "Print Listing Update Forms" under the "Listing" menu option of the CBECS Area Frame Database Application.

Input(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\SEGMENT.DBF
User input

Output(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\SEGMENT.DBF (revised)

Refer to Item G-8 in the appendix for a sample copy of the Listing Update Form.

2.2.6 Recruiting and Training Interviewers

Response Analysis field department staff recruited independent field workers in each of the 129 PSUs to staff the listing task. The number of listing segments within the PSU, the size and area of the segments, the number of buildings previously listed, and the records of the time and effort involved in listing or updating segments in the past determined the number of listers recruited in each PSU.

Listers, once recruited, were sent a set of Instructions for Fieldworkers (Item F-2 in the appendix), 1995 CBECS Training Exercise #1 (Item F-3), and all the materials needed to complete the exercise and return it to Response Analysis. Once returned, the exercise was corrected and lister strengths and weaknesses were recorded on the Fieldworker Evaluation Form (Item F-4.) The exercise was returned to the lister, marked with comments and corrections, along with Training Exercise #2 (included in Item F-6 in the appendix) and the materials needed to complete it.

Telephone training sessions were scheduled with groups of three or four listers. Listers were grouped together based on performance in Training Exercise #1. The more advanced listers were trained separately from the less advanced listers to ensure that the pace of the training was appropriate for all participants.

Training sessions were conducted by members of the Response Analysis research staff. Trainers studied the evaluation of each lister scheduled for a particular training session and conducted the training based on the strengths and weaknesses of the trainees and the loosely scripted format established prior to training. To ensure consistent training of all listers and that important points were not missed during a training session, materials were developed and a script was created to use in conducting the training. Before the lister training sessions, research staff created Training Exercise #2 and Examples for 1995 CBECS Training. A mock training was held for listing trainers using the materials mentioned above. Following the training session, the materials were modified based on input from the listing trainers and the Script for 1995 CBECS Trainers was developed to use during the training sessions. Trainers were not required to read the script verbatim during the training sessions it was used to remind the trainer of the topics that had to

be covered, to provide correct answers to examples and exercise questions, and to keep the sessions on schedule.

Trainers described the project background and objectives, went over the Examples for 1995 CBECS Training, and conducted Training Exercise #2. Training sessions were designed to be interactive. At the end of the session listers were given an opportunity to ask questions and raise concerns. All of the materials used in the telephone training can be found in Item F-6 in the appendix. Once the training was complete, trainers noted each lister's performance in the training session on the Fieldworker Evaluation Form.

2.2.7 Sendout and Updates

Once fieldworkers completed their training session and received their trainer's approval, field department staff sent them their assignments. The number of assignments sent to each lister was determined by the lister's performance on Training Exercise #1 and in the Telephone training session, the size and number of the assignments in the lister's area, and the number of listers available in a particular area. Typically, two or three segments were given to a lister to be updated as part of the listers first assignment.

Materials were prepared for each segment based on the directions provided in the Assignment Preparation Instructions (Item G-1 in the appendix.) Materials included a Listing Update Form (Item G-8), Listing Map(s) (Item G-9), an Orientation Map (Item G-10), segment sketches (Item G-5), and the required listing forms, which included Additional Building Forms (Item G-3), Additional Sketch Forms (Item G-6), Block Listing Forms (Item G-4), and Facility Listing Forms (Item G-7.) Along with their first assignment, fieldworkers were sent a letter of introduction (Item G-2) on Response Analysis' letterhead to show persons who had questions about the legitimacy of the project or the intentions of the lister. The letter, signed by David Carroll, included the Response Analysis 800 number for curious parties to call with their questions.

Listers reported their progress to Response Analysis field department staff weekly and were encouraged to call field department staff with any questions regarding procedures and irregular situations. Questions that field department staff could not answer were referred to research staff for resolution. Some listers traveled to complete assignments in areas of the country where fieldworkers could not be recruited or where the fieldworkers recruited could not complete the work in the time allowed.

■ Purpose: This menu option recounted the number of blocks in each segment. This number changed occasionally when blocks were split or combined. Field interviewer progress reports included the number of blocks in the segment.

Procedure: Select the menu option "Recount blocks in each segment" under the "Update" menu option of the CBECS Area Frame Database Application.

Input(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\BLOCK.DBF
N:\START\INTRVWRS\DATA\STUDHIST.DBF

Output(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\SEGMENT.DBF (revised)
N:\START\INTRVWRS\DATA\CBEXRDSP.DBF (revised)

- Purpose: This menu option allowed the user to enter the cost of listing each segment based on the hourly rate of the assigned interviewer and the number of hours spent on each task. Expenses were included, and were broken down by type.

Procedure: Select the menu option "Costs of Listing" under the "Listing" menu option of the CBECS Area Frame Database Application.

Input(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\COSTS.DBF
User input

Output(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\COSTS.DBF (revised)

- Purpose: This Clipper program was used by the Field Department to record the progress of field listers in completing their assignments. It presented a data entry screen to enter the number of blocks completed in each segment.

Procedure: Run the Clipper program CBEXFLD.EXE.

Input(s): N:\START\INTRVWRS\DATA\FLDINTR.DBF
N:\START\INTRVWRS\DATA\STUDHIST.DBF
N:\START\INTRVWRS\DATA\CBEXRDSP.DBF
N:\START\INTRVWRS\DATA\CBEXSDSP.DBF
N:\START\INTRVWRS\DATA\SEGMENT.DBF
User input

Output(s): N:\START\INTRVWRS\DATA\CBEXRDSP.DBF (revised)

2.2.8 Processing Materials Returned from the Field

2.2.8.1 Initial Review

Each assignment went through the initial review process shortly after it was returned from the field to ensure that all materials had been returned, to evaluate the listers performance, and to identify any major problems with the listing. The reviewer recorded problems with the listing on the Initial Review form (Item H-2 in the appendix.) General problems, as well as a brief comment about each assignment, were recorded on the Fieldworker Evaluation Form. Fieldworkers were contacted to clear up any questions or problems with the assignment and to discuss any general problems the lister was having. Fieldworkers whose work was unacceptable were removed from the study. Refer to Item H-1 in the appendix for complete instructions for the initial review.

2.2.8.2 Coding

Assignments were coded within five days of being reviewed so that problems with field procedures or fieldworkers could be caught and corrected early. Coders used materials returned from the field, information maintained in Response Analysis segment folders, and answers to questions asked of the listers during the review process.

Listing materials were prepared to simplify the keying of update information to the database following the steps enumerated in the Instructions for Coders. The purpose of coding was to make changes to the LUF easier for keyers to read as they entered information to the database, give codes when required, change building information based on information found in listing materials, fill out and code forms to add new building and facility records to the database, resolve conflicting information, and maintain the integrity of the building and facility identification numbers.

Coders flagged any block or line numbers that needed to be changed and any buildings that needed to be resurrected. Such decisions had to be checked by researchers and had to be made before keying could begin. All segments were scanned by researchers for line numbering problems and resurrected buildings before keying. Refer to the appendix for a complete set of coding instructions (Items H-3 - H-7.)

2.2.8.3 Data Entry, Proofreading, and Corrections

Once listing materials for a segment were coded, the database was updated to reflect changes and additions made to that segment in the field and during coding. Building and facility records were edited through editing windows in the database, additions were made using building and facility adding screens, and information for re-list blocks was entered through special re-list screens. Database screens were formatted to mimic the arrangement on the corresponding forms.

Keyers entered information for one segment at a time. Each segment required several steps of keying; these steps were performed in the following order to prevent the loss of information:

- 1) Block number changes, line number changes, and resurrected buildings were handled by research staff,
- 2) Information in the Listing Update Form was keyed to the database,
- 3) New buildings records were added,
- 4) New facility records were added, and
- 5) Re-list blocks were entered.

Proofing LUF's were printed for segments once keying was complete. The proofing LUF was designed to resemble the original LUF without extraneous information used only by fieldworkers. Editors compared the proofing LUF's to the original listing materials to make certain that the changes and additions made by coders and fieldworkers were made to the database and that those changes were correct. Editors followed the proofreading instructions found in Item H-8 in the appendix.

Edits recorded on the proofing LUFs were made to the database through building and facility record editing windows. Certain errors discovered during proofing could not be corrected on the proofing LUF and required flagging of the original listing materials. Examples of such edits are: additional buildings that were never added, buildings that needed to be resurrected (buildings that were mistakenly deleted in an earlier CBECS), and facilities that were mistakenly deleted. These errors were corrected using the appropriate database windows.

- Purpose: This menu option allowed the user to add new buildings in blocks that were not relisted.

Procedure: Select the menu option "Add new buildings" under the "Data" menu option of the CBECS Area Frame Database Application.

Input(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\BUILDING.DBF
User input

Output(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\BUILDING.DBF (revised)

- Purpose: This menu option allowed the user to add new facilities in blocks that were not relisted.

Procedure: Select the menu option "Add new facilities" under the "Data" menu option of the CBECS Area Frame Database Application.

Input(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\FACILITY.DBF
User input

Output(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\FACILITY.DBF (revised)

- Purpose: This menu option allowed the user to add new buildings and facilities in blocks that were relisted.

Procedure: Select the menu option "Relist buildings" under the "Data" menu option of the CBECS Area Frame Database Application.

Input(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\BUILDING.DBF
M:\START\PROJ\8800DOE\CBECS\DATABASE\FACILITY.DBF
User input

Output(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\BUILDING.DBF (revised)
M:\START\PROJ\8800DOE\CBECS\DATABASE\FACILITY.DBF (revised)

- Purpose: This menu option allowed mass updates (within a segment) of city and town information on block records. The user entered the beginning and ending blocks, and the new information. All blocks between and including the beginning and end were updated.

Procedure: Select the menu option "New P1 Cards" under the "Update" menu option of the CBECS Area Frame Database Application.

Input(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\BLOCK.DBF
User input

Output(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\BLOCK.DBF (revised)

- Purpose: This menu option printed a report showing the keying status of all segments.

Procedure: Select the menu option "Keyed Segments Reports" under the "Reports" menu option of the CBECS Area Frame Database Application.

Input(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\SEGMENT.DBF

Output(s): Printed report

- Purpose: This menu option checked all segments that had completed the keying process but had not yet been printed for proofreading. If the segment contained no blocks with problem facilities, it was tagged as ready for proofreading and in a later step a short version of the Listing Update Form would print out. Problem facilities were facilities that claimed a building in the first line that was missing or marked as ineligible.

Procedure: Select the menu option "Identify segments ready for proofing" under the "Proofing Returns" submenu under the "Reports" menu option of the CBECS Area Frame Database Application.

Input(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\BLOCK.DBF
M:\START\PROJ\8800DOE\CBECS\DATABASE\FACILITY.DBF
M:\START\PROJ\8800DOE\CBECS\DATABASE\BUILDING.DBF
M:\START\PROJ\8800DOE\CBECS\DATABASE\STATUS.DBF

Output(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\STATUS.DBF (revised)

- Purpose: This menu option printed a short version of the Listing Update Form for those segments that had completed the keying process. These were used for proofreading the changes made to the database.

Procedure: Select the menu option "Print proof-reading LUFs" under the "Proofing Returns" submenu under the "Reports" menu option of the CBECS Area Frame Database Application.

Input(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\SEGMENT.DBF
M:\START\PROJ\8800DOE\CBECS\DATABASE\BLOCK.DBF
M:\START\PROJ\8800DOE\CBECS\DATABASE\FACILITY.DBF
M:\START\PROJ\8800DOE\CBECS\DATABASE\BUILDING.DBF

Output(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\CHKNAPS\PXXXXYY.TXT, where XXX was the PSU and YY the field unit of the segment printed.

- Purpose: This menu option allowed the user to enter the status of the proofreading and correcting process for each segment.

Procedure: Select the menu option "Enter status of proofing" under the "Proofing Returns" submenu under the "Reports" menu option of the CBECS Area Frame Database Application.

Input(s): M:\START\PROJ8800DOE\CBECS\DATABASE\STATUS.DBF
User input

Output(s): M:\START\PROJ8800DOE\CBECS\DATABASE\STATUS.DBF (revised)

- Purpose: This menu option allowed the user to enter the intensity code for buildings that belonged to segments which had completed the proofing and correcting process. Only eligible buildings that were of size code 4 or 5 were presented to the user for recoding.

Procedure: Select the menu option "Code new use (intensity)" under the "Update" menu option of the CBECS Area Frame Database Application.

Input(s): M:\START\PROJ8800DOE\CBECS\DATABASE\BUILDING.DBF
User input

Output(s): M:\START\PROJ8800DOE\CBECS\DATABASE\BUILDING.DBF (revised)

- Purpose: This menu option printed a general report showing the number of segments at each stage of development during the listing and updating process.

Procedure: Select the menu option "General Progress Report" under the "Reports" menu option of the CBECS Area Frame Database Application.

Input(s): M:\START\PROJ8800DOE\CBECS\DATABASE\SEGMENT.DBF
M:\START\PROJ8800DOE\CBECS\DATABASE\STATUS.DBF

Output(s): Printed report

2.2.8.4 Consistency Checks and Facility Review

Once information for all segments was updated and entered into the database, steps were taken to improve the quality of data. The following are improvements to and checks of the database that occurred after updates were complete.

- 1) Size class (4) and (5) buildings were coded with new use codes based on building energy intensity type. The intensity codes ranged from (1) through (3). A code of (1) represented high energy intensity and included buildings used for: food sales, laboratories, and health care. A code of (2) represented medium energy intensity and included buildings used for: refrigerated storage, mercantile services, education, food service, public assembly, office space, lodging, and other. A code of (3) represented low energy intensity and included buildings used for religious worship and non-refrigerated storage. Low intensity also included vacant buildings.

The new building use for size class (4) and (5) buildings allowed a more optimal oversample of high intensity buildings. Size classes (1) through (3) were not re-coded as the new use code would have had a minimal effect with the smaller size classes and re-coding would have required more time than what remained in the listing phase.

- 2) All facility records were printed and checked. An editor scanned the facility names and descriptions to determine which facilities might be out-of-scope. Facilities likely to be out-of-scope were checked in the original listings and update materials. If an editor could determine, with certainty, that the facility was out-of-scope, the buildings in the facility were given the appropriate deletion source and deletion reason codes.
- 3) Several checks were made of facility records to ensure that buildings grouped together in a facility were treated consistently. Facility records that violated established criteria were looked-up and corrected. A check was run to make certain that buildings representing the first or last line numbers of a facility were not deleted, except with deletion reasons (7), (8), or (97). A check was run to make certain that buildings representing the first or last line numbers in facilities existed. Finally, a check was run to make certain that there were no facilities with only one building.
- 4) A database search was done of building names for the text strings: "barn," "greenhouse," "sewage," "treatment," "manu," or "mfg." Building records for buildings whose names contained any of those strings were printed and checked. Buildings determined to be out-of-scope were given the appropriate building deletion source and deletion reason codes. If the editor was unable to decide whether the building met the criteria for deletion the building was not deleted.
- 5) Buildings with no name, address, and use information were investigated. Information was placed in one of the three fields.
- 6) Buildings with no size code were investigated. If no size information could be found on listing materials, calls were made to the field lister or the facility to find out an approximate building size.

2.3 SAMPLE SELECTION

As noted in Section 1, the CBECS sample is drawn from an area frame, supplemented by list frames of large buildings. The target number of completes for 1995 was 5,500. Stratification goals called for sampling larger buildings at higher rates. The integration of the area frame and list frames, along with changes in the sample for 1995 made the sample selection procedures for 1995 more complex than those implemented for previous surveys. The area frame sample selection procedures included:

- 1) Elimination of targeted area frame cases that overlap with list frame cases
- 2) Setting final sampling rates
- 3) Selecting area frame sample cases
- 4) Elimination of targeted area frame sample cases that overlap with list frame cases

A copy of the Final Sampling Plan for the Area Frame (Item A-1) can be found in the appendix.

2.3.1 Eliminating Area Frame Cases That Overlap with the List Frame

For the 1995 CBECS, an attempt was made to reduce the overlap between the area frame and the list frames. Since large buildings and certain types of buildings are more likely to overlap, the following overlap detection procedures were used:

- 1) A search was done for the text strings "hosp," "med," "univ," and "coll" in facility names. A list of facility records containing these strings was compared to the hospital list and the college and university list to find out if the facility overlapped with the list frame. Buildings in facilities that appeared in the hospital list were given an overlap code of (1); if the facility appeared on the college and university list it was given an overlap code of (2).
- 2) Size class (5) buildings were printed and compared to the large buildings list to search for overlap. If an editor could establish with certainty that a size class (5) building appeared on the large building list, the building was given an overlap code of (5).

A total of 670 buildings were eliminated from the area frame through this process.

2.3.2 Setting Final Sampling Rates

Sample selection involved selecting a stratified sample of buildings from each of the 1995 CBECS selected area segments. The buildings within each area segment were stratified into 18 strata consisting of combinations of five square footage categories and either three energy intensity classes or four building type codes. (Stratification based on energy intensity appeared to reduce the sampling error in 1992 CBECS relative to the building type stratification conducted in prior studies. However, given that the benefits were less for smaller buildings and the costs associated with changing the classification procedures high, the decision was made to only stratify the larger square footage categories by energy intensity.) Buildings within each of the strata were selected based on a specified probability of selection determined in part through an optimal allocation study of the 1992 data. The term optimal implies that the resulting sampling rates minimized the sampling error in the 1992 CBECS total estimate of MBTUs consumed for a sample of 5,500 expected interviews. Appendix B of the Final Sampling Plan for the 1995 CBECS provides additional information on the comparison of different stratification methods, the results of the optimal allocation study, and the preparation of the initial trial sampling rates.

The results from the optimal allocation study provided us with a base for the sampling rate to apply to each stratum. However, since the optimal rates were based on the 1992 CBECS data, the differences between the 1992 and the 1995 study had to be accounted for in preparing the final rates. By reviewing the differences and similarities in the two studies, the initial optimal sampling rates were adjusted to create a set of sampling rates to yield the desired number of interviews in each stratum to total approximately 5,500. To adjust these sampling rates, the level of growth in the building population since 1992, differences in the 1992 and 1995 expected survey response rates, building eligibility differences, and differences in how the buildings were sampled in the field from the list frame facilities for which building lists were not available were considered. Also taken into consideration was how the initial estimates of the buildings's size would vary from the final reported values. After applying these factors, an additional trial and

error type procedure was used to develop the final rates through successive iterations of applying the sampling rates to the area frame database and reviewing the results.

To reach the desired number of interviews for buildings 100,000 square feet or greater, all of the size class 5 buildings in the area segment were selected. As a result, equal sampling rates were used for each energy intensity category in this size class. Furthermore, for buildings between 50,000 and 99,999 square feet, the primary activity data used to code intensity was often unknown or unreliable. Hence, while plans initially called for the use of the characteristic in the sample design for large buildings, the energy intensity categorization actually played little role in the final sample design. The target area frame selection probabilities are listed in Table 2-1 below.

Table 2-1 - Final Area Frame Sampling Rates

USAGE CATEGORIES	SQUARE FOOTAGE CATEGORIES				
	1 1K - <10K	2 10K - <25K	3 25K - <50K	4 50K - <100K	5 100K +
Type 1, Retail	0.00020	0.00239	0.01101	n/a	n/a
Type 2, Offices or <u>Unknown</u>	0.00019	0.00306	0.00703	n/a	n/a
Type 3, Open Space	0.00019	0.00232	0.00591	n/a	n/a
Type 4, Other	0.00037	0.00260	0.00686	n/a	n/a
Intensity 1, High	n/a	n/a	n/a	0.01191	0.0375
Intensity 2, Medium or <u>Unknown</u>	n/a	n/a	n/a	0.00938	0.0375
Intensity 3, Low	n/a	n/a	n/a	0.00575	0.0375

2.3.3 Selecting the Area Frame Sample

A random probability sample of buildings was selected from each sampling stratum using the sampling rates presented in Table 2-1. A description of the sample selection procedures, including an example of the programming steps is outlined in the memorandum "1995 CBECS Building Sample Selection Procedures," dated June 19, 1995. This memorandum was generalized to describe the same general selection process that was used throughout the 1995 CBECS study for the selection of area sample buildings, list frame facilities or list frame buildings. Note that within area segments, selection rates were determined by dividing the desired overall rate in Table 2-2 by the area probability of selection. In some cases the area segment probability of selection was smaller than the desired rate. As a result, the desired rate could not be obtained from these segments.

The area frame originally contained a total of 82,499 building based records. From this file a total of 14,360 records were eliminated as ineligible. From the remaining eligible 68,139 records, 670 buildings were ultimately eliminated prior to selecting the area frame sample as list frame matches to yield a file of 67,469 records. From these records, an area sample of 6,583 buildings was selected. Note that a total of 28 area sample buildings were determined to reside on the list frames after the sample selection process and were eliminated from the sample to yield a final sample of 6,555 area frame sample buildings.

Table 2-2 presents the number of listings in the area frame by square footage, the number of buildings selected in each square footage class and the average selection rate by square footage class. The table also provides the final average area sample building probability of selection by square footage.

Table 2-2 - Summary of Area Frame Sample Characteristics

Item	1K to 10K sqft.	10K to 25K sqft.	25K to 50K sqft.	50 to 100K sqft.	100K+ sqft.	Total
1995 Area Frame Listings	53,949	8,603	3,117	1,395	1,075	68,139
Area Frame to List Frame Deletions	254	114	146	75	81	670
Final Area Frame Listings	53,695	8,489	2,971	1,320	994	67,469
Area Sample Buildings Selected	1,128	1,844	1,696	991	924	6,583
Final Average Building Probability of Selection	.00026	.00256	.00761	.00974	.0156	.0064
Average Area Segment Probability of Selection	.0114	.0117	.0126	.0131	.0185	.0131
Average Within Area Segment Probability of Selection	.0262	.2372	.6461	.8128	.9760	.4967

The sample selection process was carried out in a two phases. For the first phase, a sample of buildings from all of the selected area segments was selected. Upon review of the selected records, it appeared that some of the segments needed to be corrected. A subset of the original segments were re-processed for phase two to correct the problems. Note that during the area frame-to-list-frame matching process some records were marked as duplicates erroneously. These records were fixed and the sampling process repeated. The duplicate counts presented in Table 2-1 reflect the final number of duplicates detected after all corrections were applied.

Both a main sample and a supplemental sample were selected from each area segment. The main sample contained 6,583 building records and the supplement 982 buildings. The supplemental sample was prepared to fulfill the potential need for additional sample during the field operation. The output files contained matching ID information for linking the selected files to the database and the building's probability of selection. Two probabilities of selection were provided on the file. The first one represented the building's probability of selection given that only the main sample is used and the second one represented the selection probability given both the main

sample and the supplement were used. Note that the supplemental sample records were not needed.

Listed below are the programming files involved in the sample selection process.

- Procedure: Phase 1--This SAS program reads in an ASCII file of eligible area frame buildings and selects a main sample and a supplemental sample of buildings by sampling stratum. The program computes the within area segment probability of selection based on the overall desired rate from Table 2-2 and the area segment's probability of selection. Using the within segment sampling rates, a random sample of buildings is selected from each area segment within each sampling stratum. The selected records are stored in files for future matching with the area frame database.

Input(s): M:\START\PROJ\8800DOE\CBECS\AREAFRAM\DATA\AREADOWN.TXT

Output(s): M:\START\PROJ\8800DOE\CBECS\AREAFRAM\DATA\SELECTB\AREA1.DAT
(Main)
M:\START\PROJ\8800DOE\CBECS\AREAFRAM\DATA\SELECTB\AREA2.DAT
(Supplemental)

Program: M:\START\PROJ\8800DOE\CBECS\AREAFRAM\DATA\SELECTB\AREAPULL.SAS

- Procedure: Phase 2--This SAS program repeated the process for phase 1 for a specified list of area segments. The program also created a final SAS dataset of the entire list of sampled buildings for the main sample and the supplement

Input(s): M:\START\PROJ\8800DOE\CBECS\AREAFRAM\DATA\AREAREDO.TXT.

Output(s): M:\START\PROJ\8800DOE\CBECS\AREAFRAM\DATA\SELECTB\RAREA1.DAT
(Main)
M:\START\PROJ\8800DOE\CBECS\AREAFRAM\DATA\SELECTB\RAREA2.DAT.
(Supplemental)

M:\START\PROJ\8800DOE\CBECS\AREAFRAM\DATA\SELECTB\FSAMPA1.SD2
(Main SAS dataset; 6,583 sampled records)

M:\START\PROJ\8800DOE\CBECS\AREAFRAM\DATA\SELECTB\FSAMPA2.SD2
(Supplemental SAS dataset; 982 sampled records)

M:\START\PROJ\8800DOE\CBECS\AREAFRAM\DATA\SELECTB\REDOSEGS.SAS.

2.3.4 Post Sample Elimination of Overlap of Area Frame with List Frame

Selected buildings were printed and compared against the list frame to determine if any overlap was missed in earlier checks. The lists were scanned to determine which buildings would likely appear on the list frame. Buildings that were suspected to be overlap, but were not investigated previously, were highlighted and given to an editor to check against the list frame. Examples of buildings highlighted are:

- Size class (4) buildings,
- University and college buildings,
- Hospital buildings and medical centers, and
- Government buildings.

Editors compared the buildings to the appropriate list frame lists; a building found to be overlap was given an overlap code of (A) if it appeared on the hospital list, (B) if it appeared on the university and college list, (D) if it appeared on the GSA list, or (E) if it appeared on the large buildings list. Building records remained selected, but were crossed off the selected buildings list and were not printed on the selected buildings file.

2.4 PREPARING MATERIALS FOR DATA COLLECTION

The final step in the Area Frame processing was to prepare the segments for data collection according to the instructions in the Preparation Checklist document (Item H-9 in the appendix.) This involved preparing the data collection materials and preparing the final data files.

2.4.1 Preparing Listing Update Forms

Lists of selected buildings were created for each segment with at least one selected building. The Selected Buildings List (Item I-1 in the appendix) listed all buildings in blocks with at least one selected building. Only blocks with selected buildings were listed to reduce the size of materials interviews would need for an assignment. Buildings were listed in order, around the block, to help the interviewer locate selected buildings.

Selected buildings were noted by the word "SELECTED" on the building record. All building records, selected or not, contained building line number, building name, building address, building use, building description, and building size range. Building records for buildings that were part of a multi-building facility also contained notation explaining the existence of a facility and the facility level information for that building.

For each segment, an original and three copies of the Selected Building Lists were created. The original was bound, two of the three copies were stapled along the top, and the third copy was clipped. The bound original and two stapled copies were sent to Westat and the clipped copy was kept by Response Analysis and filed with the 1995 listing materials for that appropriate segment.

2.4.2 Preparing Sketches and Maps

Sketches were prepared following the proofing instructions (Item H-8 in the appendix.) Sketches were checked to ensure that they accurately represented the information contained in the database and that they were relevant following changes made in listing procedures and criteria.

Original sketches were sent to Westat so that the interviewer would have the best possible materials when locating buildings. The two stapled copies were sent to Westat for use by the in-house and field supervisors. A set of copies was kept by Response Analysis.

2.4.3 Creating the Selected Buildings File

After all updates had been completed, the selected buildings and facilities were downloaded from the database application to ASCII text files for Westat using a menu option on the database application. Two files were created. Selected buildings that were not duplicates of buildings in the List Frame sample were downloaded to SBF_BLD.TXT, and if the building was part of a facility the facility information was downloaded to SBF_FAC.TXT.

- Purpose: This menu option creates the selected buildings files.

Procedure: Select the menu option "Selected Building/Facility File" under the "Listing" menu option of the CBECS Area Frame Database Application.

Input(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\STATUS.DBF
M:\START\PROJ\8800DOE\CBECS\DATABASE\BLOCK.DBF
M:\START\PROJ\8800DOE\CBECS\DATABASE\BUILDING.DBF

Output(s): M:\START\PROJ\8800DOE\CBECS\DATABASE\SBF_BLD.TXT
M:\START\PROJ\8800DOE\CBECS\DATABASE\SBF_FAC.TXT

CHAPTER 3

LIST FRAME SAMPLE PROCEDURES

CHAPTER 3: LIST FRAME SAMPLE PROCEDURES

The purpose of the list frame sample is to supplement the area frame sample with large buildings (i.e., buildings over 250,000 square feet.) The steps in selecting the list frame sample are:

- 1) Developing five lists of establishments that have large buildings
- 2) Selecting a sample of establishments from each list
- 3) Screening selected establishments to obtain lists of buildings
- 4) Selecting a final sample of buildings
- 5) Preparing CBECS interviewer locating materials

In comparison to the area segment listings, the list frame files are primarily facility based (or construction project based) as opposed to building based. As a result, selecting a sample of buildings from these files was a multistage process. A detailed description of each step is provided in this section. In addition, relevant reports, instructions, and materials are included in the appendices.

3.1 DEVELOPING LIST FRAME LISTINGS

Based on an analysis of data from the 1992 CBECS, five list-frame listings were targeted for development.

- 1) Government facilities
- 2) Colleges and universities
- 3) Hospitals
- 4) Other pre-1986 large buildings
- 5) Other post-1985 large buildings

The general targets for the listings were to include units that:

- 1) had buildings that were 250,000 square feet or larger, and
- 2) were located within the 129 area frame PSUs.

The individual targets for each frame listing differed, depending on the expected structure of the units in the available lists.

The 1995 special lists are described below. A summary of the total number of records in each of the files, by square footage, that were eligible for facility-level sample selection is presented in Table 3-1.

3.1.1 Government Facilities

The initial list was developed from the ASI Federal Facility Database. This database was obtained from the Federal Energy Management Program (FEMP.) The FEMP database was developed from the Government Services Administration (GSA) real property file. This file was stored in a Clipper database and contained a total of 20,288 non-DOD records. The list is facility based and each record represents all of the government-owned buildings located on the central facility. The file provides information on the number of buildings, by type, located at the facility. The following exclusion procedure limited the federal facilities list to 286 units.

<u>Step</u>	<u>Procedure</u>	<u>Buildings Remaining</u>
1	Exclude DOD facilities	20,288
2	Limit to facilities with at least 250,000 square feet in one building type	552
3	Limit to 50 states and D.C.	543
4	Limit to facilities with >500,000 square feet <u>or</u> 250,000-500,000 square feet and fewer than 10 buildings	517
5	Limit to facilities in CBECS PSUs with usable addresses	286

3.1.2 Colleges and Universities

Two initial facility based lists were obtained to create a list of college and university records. One file was obtained from the National Center for Educational Statistics (NCES) and a second from Market Data Retrieval (MDR). Details on the initial processing conducted on these two files can be found in the memorandum "CBECS List Sample Development: College File Preparation" dated April 6, 1995 (Item J-2 in the appendix.) The initial processing steps reduced and deduplicated the records to a file that contained 943 facility records with an estimated square footage of 250,000 sqft or more in the 129 PSUs. Square footage was based on total school enrollment using an estimate of 167 square feet per student. This list was later restricted to those facilities with an estimated 1,000,000 square feet or greater.¹ The final list contained a total of 455 eligible facilities.

¹Most college and university campus include a large number of buildings. Since we are targeting buildings greater than 250,000 square feet, we include only those colleges and universities with an estimated 1,000,000 square feet, with the expectation that smaller colleges and universities would be unlikely to have any buildings that are greater than 250,000 square feet. We estimated that a college or university has 167 square feet per enrolled student. We restricted the list to colleges and universities with at least 5,988 students

3.1.3 Hospitals

The initial list was obtained from the American Hospital Association (AHA). The list was restricted to hospitals in the 129 selected PSUs that were estimated to be at least 250,000 square feet (have at least 208 beds using an estimate of 1,200 sqft per bed). The final file included 1,017 list entries.

3.1.4 Pre-1986 Other Large Building List

The Large Building List (LBL) was developed in 1986, and is primarily building based with some multi-building facility records. See the **1986 NBECS Sample Selection Procedures Manual** for more information on the development of the list. The non-MSA buildings under 250,000 square feet were excluded. The final list includes general use buildings over 250,000 square feet in the 129 PSUs for a total of 6,692 list entries.

3.1.5 Post-1986 Dodge Major Projects File

This list includes three frames of projects:

- a. Built Between January 1, 1986 and December 31, 1988 (1989 Frame)
- b. Built Between January 1, 1989 and December 31, 1991 (1992 Frame)
- c. Built Between January 1, 1992 and December 31, 1994 (Note that Dodge provided the 1992 file separately from the 1993-1994 file as the corresponding data came from two separate internal Dodge sources) - (1995 Frame)

The files were obtained from McGraw Hill, Construction Information Group. The files contained a total of 2,587 project based records within the 129 PSUs.

Table 3-1: List Frame Source Files Records Eligible For Facility Level Selection By Source and Estimated Square Footage Within 129 PSUs

File Source	250,000-499,999 sqft	500,000 to 999,999 sqft	1 Million to 3,999,999 sqft	4 Million +	Total
Total Records (Facilities or Buildings)					
GSA	95	105	81	5	286
AHA	662	282	72	1	1,017
Colleges	N.A.	N.A.	455	0	455
Large (Buildings)	4,537	1,611	519	25	6,692
Dodge 1989	759	229	77	0	1,065
Dodge 1992	778	219	57	1	1,055
Dodge 1995	326	112	29	0	467
Total	7,157	2,558	1,290	32	11,037

3.1.6 List Frame Integration

The lists used to develop the list frame overlapped; a building or facility was likely to appear on more than one list. The goal was to develop a single list frame, with buildings appearing only on one of the five lists that comprise the list frame. A multistage approach was used to eliminate duplicates between lists.

To facilitate the integration of the list, a list hierarchy was established. This hierarchy was used to determine the list on which an individual building should be located. The strategy for defining the hierarchy was to put the most comprehensive lists at the top of the hierarchy. For example, the large building list targeted individual buildings, while the hospital list targeted hospital facilities. Hospitals were deleted from the large building list if the hospital was listed on the hospital facility list.

The first round of list frame integration was completed in 1986. College buildings, government buildings, and hospital buildings were deleted from the pre-1986 large buildings list. The remaining list frame integration procedures were carried out as part of the establishment sampling and screening work.

3.2 SELECTING A SAMPLE OF ESTABLISHMENTS FROM THE LIST FRAME LISTINGS

In this stage of the procedure, a sample of establishment records were selected from each of the lists. The target selection rates are listed in Table 3-2.

Table 3-2
1995 Overall Desired Sampling Rates for Facility Size Classes 6, 7, 8, and 9

Sampling Rates	Facility Size Class			
	6 250-<500K	7 500K<1M	8 1M<4M	9 4M+
1995 Facility Level Probability of Selection	.075	.15	.25	1

The actual selection rates are bounded by the selection probability associated with the PSU.

$$P_{hci} = P_h * f_{whci}$$

$$f_{whci} = \min (1, f_c/P_h)$$

where:

P_{hci} = the probability of selection of unit i in size/use class c in PSU_n

P_h = the probability of selection of PSU_n

f_{whci} = within-PSU probability of selection of unit i in size/use class c in PSU_n

f_c = the target probability of selection for units in size/use class C

For the Dodge frames for 1989 and 1992 an additional selection step was required. The first stage was actually constructed in the prior studies. For 1989, two samples were selected from the 1989 Dodge file, and in 1992, two samples for the 1992 Dodge file. The plan was to rotate these samples with each survey implementation. (There is a time value to Dodge sample, since it has information on the developer of a project, but does not necessarily have the physical locations.) Since the sampling rates were modified in the 1995 survey, the samples designated for 1995 could not simply be reused. Rather than select completely new samples, the decision was made to subsample the facilities selected in the two combined prior samples. The total number of facilities selected from each of the sources files is given in Table 3-3. Table 3-3 provides the average overall selection rate for the list sample facilities by square footage category and the average within PSU selection rates. Note the desired overall rates could not be met in all cases as some of the PSUs selection probabilities were less than the desired rate.

Table 3-3: Total Facilities Selected From Each List Frame Source Files By Source and Estimated Square Footage Within 129 PSUs

File Source	250,000-499,999 sqft	500,000 to 999,999 sqft	1 Million to 3,999,999 sqft	4 Million +	Total
GSA	9	19	25	5	58
AHA	105	66	23	1	195
Colleges	N.A.	N.A.	158	0	158
Large	621	313	154	25	1,113
Dodge 1989	72	56	21	0	149
Dodge 1992	80	44	16	0	140
Dodge 1995	44	26	6	0	76
Total	931	524	403	31	1,889
Average Overall Probability of Selection	.0719	.1445	.2358	.9114	.1407
Average <u>Within</u> PSU Probability of Selection	.4086	.3896	.4386	1.00	.4193

After the samples were selected, attempts were made to determine if the facility was selected in prior CBECS studies. For the Dodge file and the large building list, this information was available from the input files; however, for the other files, both manual and electronic searches were conducted for the 1995 selected facilities by name, street address, city name, and postal ZIP Code among the prior CBECS samples.

For the LBL, the 1986, 1989, and 1992 CB ECS ID numbers were appended to the frame file. Note that the prior CB ECS ID number was simply a sequence number based on a specified sorting criteria. Hence, the ID numbers could be recreated by simply sorting the LBL in the same order. With the ID numbers appended to the 1995 LBL sample file, the sample could be matched to the prior selected building files to determine if the facility had been selected in prior studies. This same process was not available for the lists that were replaced (i.e., the hospital list, the colleges and universities list, and the government facilities list.) The selection programs are described below. The procedure description applies to each program; program specific information is given by each type of list.

- Procedure: The SAS program reads the appropriate list source facility based file and removes any ineligible records. PSU membership and selection probability is added to the records via a FIPS state and county code matching process. Based on the PSU probability of selection, a with PSU probability of selection is determined for each sampling strata and a sample of records selected from each PSU and sampling strata combination. The program formats the files for output.

GSA Sample Selection

This sample selection was accomplished in two separate runs of the GOVNSLCT program.

Wave 1:

Input(s): M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\GOVINPUT.TXT
M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\PSU95.EXT
M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST92.SBF
M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST89.SBF
M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST96.SBF

Output(s): M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\PRESGSA.DAT

Program: M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\PGMS\GOVNSLCT.SAS
Wave 2:

Input(s): M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\NEWGOV.TXT
M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\PSU95.EXT
M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST92.SBF
M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST89.SBF
M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST96.SBF

Output(s): M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\PGS2.DAT

Program: M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\PGMS\GOV2.SAS

AHA Sample Selection

Input(s): M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\FAHA.DAT

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\PSU95.EXT

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST92.SBF

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST89.SBF

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST96.SBF

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\AHA95.ALL

Output(s): M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\PRESAHA.DAT

Program: M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\PGMS\HOSPSLCT.SAS

Combined MDR and NCES college Sample Selection

Input(s): M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\FCOLL.DAT

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\PSU95.EXT

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST92.SBF

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST89.SBF

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST96.SBF

Output(s): M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\PRESCOLL.DAT

Program: M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\PGMS\COLLSLCT.SAS

1986 Large Building Sample Selection

Input(s): M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LARGE86.DAT

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\PSU95.EXT

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST92.SBF

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST89.SBF

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST96.SBF

Output(s): M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\SAMPLEL.SD2
M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\PRESLAR.DAT
Program: M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\PGMS\LARSLCT.SAS

1995 Dodge Frames, SAS Data Set

Input(s): M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\FDOD95.SD2
Output(s): M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\PRESD95.DAT
Program: M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\PGMS\DODNEW.SAS

1989 and 1992 Dodge Files Sample Selection

Input(s): M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\D89Y89S.DAT
(Dodge 1989 Frame with 1989 Sample Flags)

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\D89Y92S.DAT
(Dodge 1989 Frame with 1992 Sample Flags, created in 1989)

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\D92Y89S.DAT
(Dodge 1989 Frame with 1992 Sample Flags, created in 1992)

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\D92Y92S.DAT
(Dodge 1992 Frame with 1992 Sample Flags)

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\D92Y95S.DAT
(Dodge 1992 Frame with 1995 Sample Flags)

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST92.SBF

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST89.SBF

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST.PRN

*Note: Both of the 1992 sample files from the 1989 frame were verified to be identical

Output(s): M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\PRESD92.DAT
(1992 Dodge Sample File)

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\PRESD89.DAT
(1989 Dodge Sample File)

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\DODGE89.DAT

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\DODGE92.DAT

Program: M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\PGMS\DODOLD2.SAS

- Procedure: Appends the 1986, 1989 and 1992 CBECs ID number to the selected 1995 Large building list records and identifies cases selected in the prior three studies

Input: M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LARGE86.DAT
(1986 Large Building List (LBL) Records)

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\SAMPLEL.SD2
(1995 Sampled LBL Records)

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST86.SBF
(1986 NBECS Selected Building File)

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST89.SBF
(1989 CBECs Selected Building File)

M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST92.SBF
(1992 CBECs Selected Building File)

Output: M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\RPRESLAR.DAT
(1995 Final Large Building File Sample)

Program: M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\PGMS\MATCH.SAS

- Procedure: This Clipper program assigned eligibility codes to the government list of facilities. Those that were eligible were then sent for facility-level sampling.

Input(s): M:\START\PROJ\8800DOE\CBECS\LISTDB\ORIGINAL.DBF
(from Federal Buildings Supplemental Survey)

Output(s): M:\START\PROJ\8800DOE\CBECS\LISTDB\GOVFRAME.DBF

Program: M:\START\PROJ\8800DOE\CBECS\LISTDB\GROUPS.EXE

- Procedure: This Clipper program read the files of selected facilities and added them to the CBECs List Frame database application.

Input(s): M:\START\PROJ\8800DOE\CBECS\LISTDB\PRESGSA.DAT
M:\START\PROJ\8800DOE\CBECS\LISTDB\PRES COLL.DAT
M:\START\PROJ\8800DOE\CBECS\LISTDB\PRESAHA.DAT
M:\START\PROJ\8800DOE\CBECS\LISTDB\PRESLAR.DAT
M:\START\PROJ\8800DOE\CBECS\LISTDB\PRES D89.DAT
M:\START\PROJ\8800DOE\CBECS\LISTDB\PRES D92.DAT
M:\START\PROJ\8800DOE\CBECS\LISTDB\PRES D95.DAT

Output(s): M:\START\PROJ\8800DOE\CBECS\LISTDB\FACILITY.DBF

Program: M:\START\PROJ\8800DOE\CBECS\LISTDB\SAMPLE.EXE

The FACILITY.DBF file was the basis for the CBECS List Frame database. This database held facility-level and building-level information for all selected facilities. Refer to the "1995 CBECS List Frame Procedures" document in the appendix for a complete description of the creation of the database and the structure of the database.

3.3 SCREENING SELECTED ESTABLISHMENTS

In this stage of the sample development procedures, the selected establishments were screened to obtain additional information. The process included:

- 1) Checking sampled establishments for membership on a higher-ranking list
- 2) Preparing screening materials
- 3) Screening establishments by telephone
- 4) Coding screened establishments
- 5) Updating the list frame database

It is important to note that not all selected establishments could be screened. Unscreened establishments were screened in the field as part of the CBECS interviewing procedures.

3.3.1 Removing Duplicate Records

The first procedure implemented to eliminate duplicate records was a sample-to-list comparison procedure. In this procedure, within each PSU each sampled establishment was manually compared to the establishments listed on each of the higher-ranking lists.

The second stage of overlap analysis was carried out as part of the screening process. As noted below, the screening instrument contained questions to identify the presence of an establishment on a higher ranking list.

3.3.2 Preparing Screening Materials

Screening materials consisted of a call record sheet (Item K-2 in the appendix), screener questions (Items K-3 - K-7), and a listing form for listing the buildings within a facility (Item K-9 - K-10).

Call record sheets were made for each of the facilities selected for screening. Printed on the sheets were the name and address of the facility along with the PSU and the FIPS county code of the facility. A contact name and phone number were also included when available. The different colors of the call record sheets represented each of the four frames. Government was printed on gold, colleges on blue, hospitals on gray, and large buildings on green paper.

- Procedure: Run the Report Writer report "Facility Screening" in the report library CBECS.RP1. This prints customized facility information onto pre-printed call record sheet templates.

Input(s): M:\START\PROJ\8800DOE\CBECS\LISTDB\FACILITY.DBF

Output(s): Call record sheets

Program: Report Writer RR.EXE

Screening questions were customized to the type of facility being screened, and for the most part were aimed at removing overlap from within the frames. In addition, the Large Building screener (Item K-7) was used to screen out establishments whose primary purpose is manufacturing.

All materials for a particular facility were kept in a folder. Two sets of files were produced for each piece of sample--one set with red labels for RAC copies of information and another set with blue labels for Westat. Both sets had the facility's name and RAC ID printed on them. The RAC files were sent to the phone center and each file was sent back when screening was complete. Westat files were kept in-house. When the RAC folders were returned from the phone center, any information that might prove useful for Westat was copied and put into the Westat file. This information included maps and sketches of the facility, and building lists provided by the facility that were not on the RAC building listing forms. Such information would help the respondent determine which building was being referred to during the main interview and help the interviewer prove the validity of the study. These folders were maintained throughout the remainder of the study; the final version of each folder with information relevant to data collection was sent to Westat for the data collection phase.

3.3.3 Conducting the Screening Phase

Prior to screening, facilities in the government frame were compared to the 1993 Federal Buildings Supplemental Survey. In cases where duplicates were found, if the FBSS information included a complete listing of the facility's buildings along with their uses, the case was considered complete. If there was not a complete list of buildings, it was sent to the phone center to complete the screening.

Interviewer training was conducted at the telephone center by the CBECS project managers. FBSS demonstrated that compiling building lists did not allow the interviewer to follow a questionnaire because he or she was constantly being shuffled around from contact to contact; therefore, no formal questionnaire was developed. The interviewers were given a list of possible contacts and answers to questions the respondent might have. Otherwise the interviewers were allowed to follow the respondent's lead when trying to obtain a building list. The screener was the only formally written piece and was read verbatim before trying to obtain a building list.

As part of the training, interviewers were given different scenarios that they may run into while screening the facilities, such as colleges with off-campus buildings or large buildings with many businesses in them. Techniques for getting past difficult respondents and for finding hard to achieve phone numbers were also discussed. Interviewers who worked on FBSS offered many helpful suggestions to those who were new to collecting building information.

Interviewers attempted to secure all necessary information over the phone. In some cases, the respondents preferred to submit the information on hard copy. The interviewer then faxed a cover letter, a sheet explaining the buildings eligible for CBECS, and building listing forms to the respondent. The respondent completed the forms (or sent the list in a format that was more convenient) and faxed them back to Response Analysis along with any maps or other helpful information. In a few cases, the respondents preferred to receive and return the materials by mail rather than by fax.

Because the Dodge sample is a supplement to the Large Building frame, the comparison between these two samples was intended to find errors that may have been made during screening. The Large Building list should only include buildings built before 1986. Any buildings constructed after 1986 were supplemented by the Dodge Reports. In a few cases, the respondent reported that Large Buildings were built after 1986. These cases were flagged for further review by the project staff. It was necessary to do some investigating to determine why buildings on the large building list were reported as being built after 1986 and may have required another call to the respondent. In most cases it was found that either the respondent answered the question wrong or a building was torn down and replaced by a new building which also appeared on the Dodge Report.

3.3.4 Processing Screened Facilities

3.3.4.1 Coding the Screening Status

Each facility was given a screening status based upon whether the screening was complete and whether or not the facility was eligible for CBECS and so should be included during selection. The screening statuses used were:

- 1 Complete and eligible
- 2 Incomplete--Refused
- 3 Incomplete--Cannot Locate
- 4 Ineligible--Manufacturing
- 5 Ineligible--Other
- 6 Ineligible--Duplicate on a Higher Frame
- 7 Ineligible--Duplicate on the Same Frame
- 8 Does not Exist
- A Information received too late

The following are statuses that may require further explanation.

Status 2, Refused, was used for all incomplete cases where the contact was spoken to, but no information was received. This includes both hard refusals and cases where the respondent seemed agreeable, but was not able to send a list in the limited amount of time.

Status 3, Cannot Locate, was used for all cases where no contact was made, usually because a current phone number could not be found.

Status 5, Ineligible--Other, included facilities out of scope for reasons other than manufacturing. Examples of such reasons are: a facility that is over 75% residential, a facility where over 75% of the floor space is used for parking, and a DOD facility inaccessible to non-DOD personnel.

Status 8, Does Not Exist, was used only in cases where it was confirmed that the building does not exist, either because it was torn down, or, as in the case of some Dodge sample, the building was never built.

Status A, Received too late, was used when information was received by mail from the respondent too late to be included in the selection process. As a result each of these cases was treated as a refusal during selection. All of the materials received from the facility were copied and sent to Westat.

3.3.4.2 Coding Building Records

Each building in a facility was given codes for size, use, and intensity. The codes for use and intensity were the same as were used for the area frame. The coding for both use and intensity were both three point scales. For intensity, "1" was the code for the highest intensity and "3" the code for the lowest intensity. For use, "1" was the code for a building with little open space and "3" the code for a building with much open space.

For size, the first four codes were the same as in the area frame. The fifth code (100,000 sq ft or more) was broken up into five separate codes for the list frame. These additional codes were:

5	100,000 to 249,999 sq ft
6	250,000 to 499,999 sq ft
7	500,000 to 999,999 sq ft
8	1,000,000 to 3,999,999 sq ft
9	4,000,000 sq ft or more

3.3.5 Updating the Database

While telephone interviewers were screening the facilities, they manually edited the call record sheets to reflect any changes to the facility information printed there. After the facility had either been successfully screened or had reached some other final status, the call record sheet was used to update the CBECS List Frame database. The screening status of the facility was coded from the call record sheet (see section 3.3.4.1) and entered so that staff could track the progress of the screening activity. Any new or changed information about the facility was also entered, including the address of the facility and names and addresses of contact people.

The building information was entered for those facilities that were successfully screened. The information was usually received as a printed list, although sometimes it was given over the telephone. In any event, the data were entered into the database as new buildings. Each building was linked to the facility by a unique four-digit facility ID number. Each building was also given its own unique ID number, beginning with number 1.

- Purpose: This menu option allows the user to edit facility information.

Procedure: Select the menu option "Facilities" under the "Data" menu option of the CBECS List Frame Database Application.

Input(s): M:\START\PROJ\8800DOE\CBECS\LISTDB\FACILITY.DBF
User input

Output(s): M:\START\PROJ\8800DOE\CBECS\LISTDB\FACILITY.DBF (revised)

- Purpose: This menu option allows the user to edit building information.

Procedure: Select the menu option "Buildings" under the "Data" menu option of the CBECS List Frame Database Application.

Input(s): M:\START\PROJ\8800DOE\CBECS\LISTDB\BUILDING.DBF
User input

Output(s): M:\START\PROJ\8800DOE\CBECS\LISTDB\BUILDING.DBF (revised)

3.4 LIST FRAME BUILDING-LEVEL SAMPLE SELECTION

In the prior processing steps, a total of 1,066 of the 1,889 selected facilities completed the screening and were eligible. A list of buildings was obtained for all of those facilities. In total, we had to send 407 facilities to the field for further processing that did not complete the screening, but were designated as eligible. These included a total of 182 refusals and 225 incomplete screenings. A total of 386 facilities were found to be ineligible, and 30 were found to be duplicates on other lists. The 1,066 screened facilities yielded a list of 8,613 buildings. From this list of 8,613 buildings, RAC selected a sample of 1,041 building records for the main sample.

Table 3-4 summarizes the processing steps on the facility level list frame samples.

From each screened facility (1,066) a probability sample of buildings was selected within each of the 22 sampling strata presented in Table 1-1, section 2.3.2, using the overall desired probabilities of selection presented in the table. To determine the within facility and sampling stratum probability of selection, we divided the overall desired rate from Table 1-1 for the associated stratum by the facility's probability of selection. A description of the building sample selection procedures, including an example of the programming steps is outlined in the memorandum "1995 CBECS Building Sample Selection Procedures" dated June 19, 1995 (Item A-3 in the appendix). As indicated in section 2.3.3, this memorandum gives a generalization of the selection procedures that we applied throughout the different steps of the area and list frame selection process including the area frame buildings, the list frame facilities, and the list frame buildings.

Table 3-4 Summary of the List Frame Processing Steps and Sample Selection

List Frame Source	Total Sample Facilities	Total Facilities Screened And Eligible	Total Facilities Eligible But Not Screened (Buildings Processed in the Field)	Total Facilities Determined to be Duplicates on other Lists	Total Facilities Ineligible	Total Buildings Listed among Screened Facilities (Facilities: 1,066)									
						Total Buildings Selected By Square Footage Class									
						1-10K	10-25K	25-50K	50-100K	100-250K	250-500K	500K-1M	1M+	Total	
GSA	58	52	6	0	0	133	310	158	121	88	21	19	15	2,065	
AHA	195	141	43	11	0	456	213	166	165	134	70	48	14	1,266	
College	158	121	30	7	0	1,528	666	697	683	341	47	14	2	3,978	
Large Building List	1,113	510	301	4	298	116	85	56	55	111	144	135	102	804	
Dodge 1989	149	94	20	4	31	0	0	8	9	25	38	27	17	141	
Dodge 1992	140	95	6	3	36	17	21	14	21	40	25	31	10	178	
Dodge 1995	76	53	1	1	21	0	1	3	3	17	21	30	10	85	
TOTAL	1,889	1,066	407	30	386	3,497	1,335	1,119	1,080	765	369	284	164	8,613	
Average Probability of Selection						00026	00278	0069	0086	0375	0743	1304	2136		
Average Within Facility Probability of Selection						00197	0248	0689	0638	3659	8720	9568	9964		

We conducted the list frame building selection process in waves as the screenings were completed by RAC staff. From the screened facilities, we prepared a total of two interim and one final file of the building records associated with the successfully screened and eligible selected facilities. The three input files and the corresponding three selection programs are given below.

- Procedure: SAS Program to read ASCII input file of listed buildings from selected and screened list frame facilities. Program selects a probability sample of buildings within the selected facilities and outputs the records to a specified ASCII format for subsequent merge into the Clipper building database.

Inputs: Wave 1:
M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LISTD_CO.TXT

Wave 2:
M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LISTD_C2.TXT

Wave 3:
M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LISTD_C3.TXT

Combined File (Waves 1-3):
M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\FULLLIST.TXT

Outputs: Wave 1 Main Sample Cases, 876 records
M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST1P1.DAT

Wave 1 Supplemental Sample Cases, 72 Records
M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST21P1.DAT

Wave 2 Main Sample Cases, 80 records
M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST1P2.DAT

Wave 2 Supplemental Sample Cases, 7 Records
M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST21P2.DAT

Wave 3 Main Sample Cases, 71 records
M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST1P3.DAT

Wave 3 Supplemental Sample Cases, 6 Records
M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\LIST21P3.DAT

Final SAS Database of entire set of main sample selects, 1,027 records
M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\FULL1.SD2

Final SAS Database of entire set of Supplemental sample selects, 85 records
M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\DATA\FULLA.SD2

Program: Wave 1:
M:\START\PROJ\8800DOE\CBECS\DOCUMENT\LISTFRAM\PGMS\LISTPULL.SAS

Procedure: Select the menu option "Primary Sample" under the submenu "Selected Building List" under the "Listing" menu option of the CBECS List Frame Database Application.

Input(s): M:\START\PROJ\8800DOE\CBECS\LISTDB\FACILITY.DBF
M:\START\PROJ\8800DOE\CBECS\LISTDB\BUILDING.DBF

Output(s): B1_XXXX.TXT (primary sample)
B2_XXXX.TXT (secondary sample)
(XXXX is the unique four-digit facility ID number)

After all updates had been completed, the selected buildings and facilities were downloaded from the database application to ASCII text files for Westat using a menu option on the database application. Two files were created. The first file, SBF_BLD.TXT, contained building and facility information. Criteria for inclusion in this file were: 1) the facility completed screening and the building was selected, or 2) the facility did not complete screening. The second file, SBF_FAC.TXT, contained information on facilities without building information.

- Purpose: This menu option creates the selected buildings files.

Procedure: Select the menu option "Selected Building/Facility File" under the "Listing" menu option of the CBECS List Frame Database Application.

Input(s): M:\START\PROJ\8800DOE\CBECS\LISTDB\STATUS.DBF
M:\START\PROJ\8800DOE\CBECS\LISTDB\FACILITY.DBF
M:\START\PROJ\8800DOE\CBECS\LISTDB\BUILDING.DBF

Output(s): M:\START\PROJ\8800DOE\CBECS\LISTDB\SBF_BLD.TXT
M:\START\PROJ\8800DOE\CBECS\LISTDB\SBF_FAC.TXT