6/29/98

1999 CBECS Sample Plan

I. INTRODUCTION

As specified in the modification for Task #97-006B, dated 4/16/98, the 1999 CBECS will target about 5,000 completed interviews; including about 4,500 interviews from cases that were selected for the 1995 CBECS and about 500 interviews from cases selected from an F.W. Dodge list of new construction. In addition, two pretests will be conducted to test procedures for the 1999 CBECS. Each pretest will include about 100 interviews.

The purpose of this document is to furnish preliminary recommendations regarding the procedures for developing the sampling frames and for selecting the samples for the 1999 CBECS and its pilots. In this document, we furnish information on the proposed sample frames, make recommendations regarding sample frame development procedures, and make recommendations regarding sample selection for the 1999 CBECS.

In the first section of this plan, we specify procedures for selecting a sample of buildings that were selected for the 1995 CBECS. In the second and third sections, we specify procedures for developing the F.W. Dodge new construction sample frame and for selecting a sample from this frame. In the final section, we specify procedures for selecting samples for the pretest interviews.

II. DEVELOPING THE LONGITUDINAL SAMPLE

The first sample frame consists of "buildings that were selected for the 1995 CBECS." The sample design specifications indicate that the sample frame shall consist of "both responding and nonresponding buildings." The sample design specifications do not comment on whether buildings that were determined to be "out-of-scope" and/or buildings that were determined to be "not eligible for interview" should be included in the sample frame. The sample specifications

call for 4,500 completed interviews. The sample specifications target an 80% response rate for the building survey.

A. History of the Sample Frame

The existing CBECS sample frame was first developed in 1986 for the 1986 NBECS. Subsequently, the sample frame was updated and used for the 1989 NBECS, 1992 CBECS, and 1995 CBECS.

The sample frame consists of a multistage area probability sample frame supplemented by list frames of large and special use buildings. The area frame includes 509 segments distributed over 129 Primary Sampling Units (PSUs). The purpose of the area frame is to furnish complete coverage of the population of commercial buildings. The supplemental list frames include lists of government buildings, colleges and universities, hospitals, and large buildings. These lists are restricted to the geographic area covered by the 129 PSUs. The purpose of the list frames is to increase the number of large buildings in the sample.

A longitudinal sample was used for the 1992 CBECS; a subsample of cases in the 1986 CBECS sample was selected for reinterview in the 1992 CBECS. Each sampled case was recontacted, even if it could reasonably be expected to be "out-of-scope" or "ineligible" for the 1992 survey. The procedure used to develop the 1992 sample frame is a statistically conservative but costly procedure for developing the sample frame. In addition, it is important to note that the 1992 interviews were completed through in-person contacts at the building.

B. Criteria for Inclusion in the 1999 CBECS Sample Frame

There were 8,074 sample cases selected for the 1995 CBECS (after field additions and subsampling). Of these, 6,590 were determined to be in-scope and eligible for interviews. Interviews were completed with 5,766 buildings. Buildings that were used for noncommercial purposes, buildings of 1,000 square feet or less, parking garages, and buildings that were part of a manufacturing establishment were "not eligible" for interview (1,104 buildings). In addition, structures that do not meet the definition of a building (dilapidated, demolished, or under

construction), duplicates, and out of segment buildings were considered to be "out-of-scope" (380 buildings).

Based on prior discussions between EIA and Westat, the 1999 sample frame will include responding and nonresponding buildings that were eligible for and in-scope for the survey in 1995. In addition, 21 buildings that were under construction at the time of the 1995 interview will be included in the sample frame. These buildings have a high probability of having been completed and thus are now "in-scope" for the survey. As we will discuss in the next section, these buildings are not part of the population covered by the Dodge frame.

Excluding ineligible buildings from the 1995 survey is not statistically conservative. We certainly expect that some buildings will have been changed in ways that make them eligible for the CBECS interview. However, the costs associated with rescreening ineligible buildings would be significant. Moreover, these costs were not included in the 1999 CBECS cost estimates furnished to EIA.

C. Preliminary Sampling Rates for the Longitudinal Sample

The sample design specifications in the task modification do not include explicit instructions regarding the sample rates for the longitudinal sample. It merely states that the sample from these buildings will be "slanted toward larger buildings." Since that is consistent with the sampling rates used in the 1995 survey, we assume that the sampling rates for the 1999 survey should retain the same structure as the sampling rates for the 1995 survey.

We have identified two options for setting the subsampling rates for the longitudinal sample. In the first, subsampling rates would be set to reduce the probability of selection by the same ratio for all sample cases. In the second, the subsampling rates would take into account unrealized sampling rates from the 1995 sample. The following discusses these two options in greater depth.

In addition, EIA has expressed a desire to designate certain buildings that have a significant influence on the sample estimates as certainty buildings, in order to ensure that their influence does not further increase as a result of this subsampling activity. The following specifies a procedure achieving this goal.

C.1. Uniform Proportional Subsampling Rates

Each case sampled for the 1995 CBECS had an explicit probability of selection. Those selection probabilities varied by sample frame, size class, and building type. The selection probabilities were designed to yield 5,500 completed interviews. The final number of completed interviews was 5,766. One approach to achieving a final sample of 4,500 cases would be to establish a uniform subsampling rate that would give each sampled case from 1995 the same "within frame" probability of selection.

Implementation of a uniform reduction in the sampling rate would require the estimation of two subsampling factors.

- *Eligibility factor:* The eligibility, *e,* is defined as the proportion of the 1995 CBECS in-scope and eligible buildings that are in-scope and eligible for the 1999 CBECS. In this preliminary sampling plan, we will assume that $e = 0.95$. The final estimate will be informed by the pretests and by other analyses of CBECS frame data. In the final sampling plan, we expect that the eligibility factor, *e*, may vary by type and size of building.

- *Response rate*: The response rate, *r,* is the targeted 1999 response rate, 80%.

The targeted sample size, *t*, is a function of the eligibility factor, *e*, the response rate, *r,* and the final number of interviews, *i.*

$$t = i / (e*r) = 4{,}500 / (.95*.80) = 5{,}921$$

The uniform subsampling rate, *u*, is the ratio of the targeted sample size, *t*, to the number of cases available in the sample frame, *f*.

$$u = t / f = 5{,}921 / 6{,}611 = 0.8956$$

Using this subsampling rate, approximately 90% of the cases selected for the 1995 sample would be reselected for the 1999 longitudinal sample.

C.2 Modified Proportional Subsample Rates

In the 1995 CBECS sample, there were targeted and actual sampling rates. The targeted sampling rates were set based on an optimal allocation analysis that examined the contribution of different components of the sample to the variance of sample estimates. However, in setting the final sample rates for an individual case, the actual sample rates were sometimes bounded by the selection probability for the PSU or the area segment. For example, the target sampling rate for buildings in size class five (100,000 to less than 250,000) was 0.0375. However, the actual average probability of selection was .0185. The achieved selection rates were bounded by the actual segment selection rates, which for most segments was less than 0.0375.

In addition, some buildings may have been classified in one size or type stratum at the time they were originally listed and found to be in a different stratum at the time they were interviewed in 1995. To select the longitudinal subsample, size and type information available in the 1995 interview (or in detailed interviewer observations in the case of nonresponding buildings) will be used for stratification purposes (not the original listing information). In theory, this permits a more nearly optimal allocation of the longitudinal sample to the various sampling strata. However, in practice, the benefit will be somewhat limited given the high overall rate at which the 1995 CBECS buildings will be retained for the 1999 survey.

In order to improve the sample allocation for the 1999 CBECS, one could apply the subsampling factor, *u,* to the *targeted* 1995 sampling rate, *t95,* of the sampled case, rather than to the *actual* 1995 sampling rate, *a95*. The final subsampling rate, *uf,* would be equal to the smaller of *u*t95/a95* and 1.

As before, the targeted sample size, *t*, is a function of the eligibility factor, *e*, the response rate, *r*, and the final number of interviews, *i*.

$$t = i / (e*r) = 4{,}500 / (.95*.80) = 5{,}921$$

The uniform subsampling rate, *u*, is the ratio of the targeted sample size, *t*, to the number of cases available in the sample frame, *f*.

$$u = t / f = 5{,}921 / 6{,}611 = 0.8956$$

The final subsampling rate is developed using an iterative process that tests the "take" from alternative subsampling rates. We continue to increase the subsampling rate until the 1999 subsample size, *s99*, equals the targeted sample size, *t*.

$$uf = \min(u*t95/a95, 1)$$

$$s99 = \text{Sum of } uf \text{ for all cases}$$

$$\text{if } s99 < t, \text{ increase } uf$$

Where $t95 = a95$, $uf = u$. Where $t95 > a95/u$, $uf = 1$. Where $a95/u > t95 > a95$, $u < uf < 1$.

C.3. Assigning Subsampling Rates for Influential Buildings

In the 1995 CBECS, certain sample cases had weighted estimates for energy consumption that were large enough to exert a significant influence on the final sample estimates. In particular, there were four buildings with weighted electricity consumption and six buildings with weighted natural gas consumption greater than .5% of the population total. For these buildings, the subsampling factor, *uf*, will be set to 1.

D. Summary and Recommendations

The longitudinal sample for the 1999 CBECS would include all eligible buildings from the 1995 survey, including both responding and nonresponding buildings. In addition, we recommend including the 21 cases from the 1995 survey where the building was under construction. This

procedure will fail to represent buildings that changed their eligibility status from "ineligible" or "out-of-scope" to "eligible and in-scope." However, we expect that the number of these cases will be small.

In selecting the sample, we recommend that EIA adopt the modified proportional subsampling procedure. It does not require a reanalysis of the optimum allocation for the sample. At the same time, it offers a potential improvement over a uniform proportional subsampling procedure. Our preliminary estimates are that the average sampling rate for buildings in the longitudinal sample will be about 90%.

As requested by EIA, we will set the subsampling rate to 1 for the ten buildings where weighted electricity or gas consumption is greater than 0.5% of the population total.

III. DEVELOPING F.W. DODGE SAMPLE FRAME

The second sample frame consists of "new large buildings (over 10,000 square feet) selected from the F.W. Dodge list of new construction." 500 completed interviews are targeted from this sample frame.[1] The new large buildings to be selected from the F.W. Dodge list are those that have been built since the area frame listings were made for the 1995 CBECS (discussed further below). These buildings will be sampled only from the 129 PSUs used for the 1995 CBECS.

F.W. Dodge has developed a national database of residential, commercial, and industrial construction projects. The primary users of the database are companies that have an interest in selling products or services to building developers. F.W. Dodge analysts report that coverage studies conducted by the Bureau of the Census confirm that F.W. Dodge covers at least 95% of all new commercial building construction projects. The purpose of the F.W. Dodge list is to

[1] With 500 completed interviews, the F.W. Dodge sample will "oversample" new buildings as compared with the longitudinal sample. An alternative strategy would be to select the longitudinal sample and the F.W. Dodge sample at the same rates. The "oversample" of new buildings decreases variances for estimates for new buildings, but increases variances for estimates for the population of buildings.

furnish a sample frame of commercial buildings constructed since the CBECS sample frame was updated in early 1995.

A. Previous Use of F.W. Dodge Sample Frames

F.W. Dodge lists were obtained in 1989, 1992, and 1995 to update the sample frames for those surveys.

In 1989 and 1992, the F.W. Dodge lists included a census of large construction projects (greater than 250,000 square feet) and a sample of moderate-sized construction projects (50,000 to 250,000 square feet). The F.W. Dodge list of large construction projects updated the large building list on the list frame. The F.W. Dodge list of moderate-sized construction projects increased the number of new buildings between 50,000 and 250,000 square feet in the sample and thereby ensured that there would be an adequate number of this type of buildings in the sample. In the 1989 CBECS, a listing update of 40% of the area frame sample segments also contributed new buildings to the sample. In the 1992 CBECS, a listing update of the 1989 update segments, as well as a rescreening of the list frame cases with four or more buildings also contributed new buildings to the sample.

In 1995, the F.W. Dodge list included only projects over 250,000 square feet. In addition to the update furnished by the F.W. Dodge list, the entire area frame sample was updated and all of the list frames, except for the large building list frame, were updated.

B. Proposed Structure of the F.W. Dodge Sample Frame

In the 1999 CBECS, the F.W. Dodge list is proposed as the ***only*** source of new buildings. The F.W. Dodge list is proposed to cover buildings as small as 10,000 square feet that were constructed in 1995 or later.

For the 1989 and 1992 CBECS samples, the F.W. Dodge lists were used to develop two frames. A complete list of projects over 250,000 square feet and a sample of projects over 50,000 square

feet. For the defined geographic area, there were relatively few projects over 250,000 square feet (less than 500), while there were over 15,000 projects in the 50,000 to 250,000 square foot range.

Because of the differences in these two frames (i.e., having a true frame for large buildings and a "conceptual frame" for the moderate-sized buildings) the overlap detection procedures used in the weighting process were different. Direct comparisons were made between the Dodge large building sample frame and the other frames to detect and apply the weighting solution to overlap cases. For the moderate-sized building frame, the frame was assumed to be complete. Therefore any new building from another frame that fell into the 50,000 to 250,000 square foot size group was assumed to be an overlap with the Dodge frame.

For the 1999 survey, the F.W. Dodge frame is designed to be mutually exclusive of the longitudinal frame (i.e., it does not intentionally overlap the existing frame). However, as noted in the next section, there is still an issue of overlap that must be addressed. Thus, within the limits of the budget allocated to purchase of the F.W. Dodge list, we propose to have two F.W. Dodge frames; a complete list of new construction projects over 250,000 square feet and a sample of new construction projects of 10,000 to 250,000 square feet.

C. Procedures for Developing the Sample Frame

The F.W. Dodge list will be a list of construction projects. The list of projects will be transformed into a sample frame of buildings using the following procedures:

- *Geographic restriction of projects:* We will obtain a list of construction projects that are in the CBECS PSUs. As noted above, the list will include all projects over a target size and a sample of projects below that size.

- *Selection of facilities:* We will select a sample of facilities for screening. The screening rates (covered in the next section) will be adequate to ensure that targeted sampling rates for individual buildings can be realized.

- *Screening of facilities:* We will screen the selected facilities and obtain a list of buildings eligible for the sample. The final sample of buildings will be selected from this sample frame.

When screening individual buildings included in sampled construction projects, one must first determine whether the building is eligible for the CBECS sample. In addition to the criteria used in previous CBECS surveys, there are two additional issues that must be resolved for the 1999 CBECS. The first is the treatment of building renovations and additions. The second is the cutoff date for inclusion in the sample.

- *Building renovations and additions:* The F.W. Dodge sample includes a range of project types, including new construction, building additions, and building renovations. The square footage indicator in the F.W. Dodge database indicates the floorspace covered by the project, not the amount of new floorspace. For purposes of the 1999 CBECS, the definition of a "new" building is one which was completed after the cutoff date. In the event that the Dodge record refers to an addition, the building will be treated as a "new" building only if 1) groundbreaking for the addition was after the relevant cutoff date and 2) the addition more than doubled the size of the building.

  Earlier we noted that the longitudinal sample would exclude buildings that were part of the 1995 CBECS sample frame but that had been renovated so that the building use changed from a noncommercial use to a commercial use. Conceivably, the F.W. Dodge lists could furnish a sample of buildings whose use changed from a noncommercial use to a commercial use. However, that is outside the scope of the current specifications and will not be considered further here.

- *Cutoff date:* Since the F.W. Dodge sample frame is being treated as a complete frame, not just a supplemental sample, it is important to consider how the dates covered by the longitudinal sample compare to the dates covered by the F.W. Dodge sample.

The end date for the longitudinal sample conceptually coincides with the end date for the area frame listings. Since each of the 509 listing updates was completed at a different time, there really is no specific end date. The listing updates were completed between 2/15/95 and 5/15/95. However, a good approximation is that the area frame includes buildings where the foundation was complete by 4/1/95.

By comparison, F.W. Dodge dates construction projects in their list as the date of ground-breaking for the first building in the project. Therefore, if a project were dated 1/15/95, it implies that at least one building in the project had ground-breaking occur on 1/15/95.

In previous CBECS samples, the F.W. Dodge supplemental frame was selected to include all buildings up to 12/31 of the year preceding the survey (e.g., for the 1995 CBECS, the F.W. Dodge frame included projects through 12/31/94). It is clear that if those sample procedures were applied for the 1999 sample, it would result in a conceptual overlap between the longitudinal sample and the F.W. Dodge sample. (For example, a building whose ground-breaking occurred on 2/15/95 could appear in both the longitudinal frame and the F.W. Dodge frame.)

The purpose of the frame of moderate-sized Dodge buildings is to furnish a complete frame of new buildings in this size class. As such, the frame should "dovetail" with the previous sample frame. Since the existing CBECS frame includes buildings where "groundbreaking" occurred on or about 4/1/95, the moderate-size Dodge building frame should include projects whose groundbreaking data was 4/1/95 or later. Any longitudinal frame building in this size class that was constructed in 1995 will be assumed to have had ground breaking prior to 4/1/95 and thus will not overlap with the F.W. Dodge frame.

The purpose for the frame of large Dodge buildings is to furnish a frame of new

buildings, as well as a supplemental sample of large buildings. As such, the frame should "dovetail" with the previous F.W. Dodge frame and should include projects where groundbreaking occurred 1/1/95 or later. Any longitudinal frame building that was constructed in 1995 or later will be checked for inclusion on the F.W. Dodge frame. If it does, the building will be assigned a PSU weight and will be deleted from the F.W. Dodge frame.

The recommended procedures for frame development are:

1) An F.W. Dodge project that refers to an addition will be included in the sample frame only if 1) groundbreaking for the addition was after the relevant cutoff date and 2) the addition more than doubled the size of the building.
2) The new large building frame will include projects of 250,000 square feet or more, dated 1/1/95 through 12/31/98. In the event that screening identifies a building of less than 250,000 square feet, the cutoff date is moved from 1/1/95 to 4/1/95.
3) The new moderate sized building frame will include projects 10,000 to 250,000 square feet, dated 4/1/95 through 12/31/95. In the event that screening identifies a building of less than 10,000 square feet, the building is deleted from the frame.
4) An overlap analysis will be conducted for buildings over 250,000 square feet in the longitudinal frame that had a construction year of 1995. If the building is found in the F.W. Dodge new large building frame, it will be deleted from the frame and will receive a PSU weight in the longitudinal sample. (Note: The longitudinal frame includes both area frame and list frame buildings from the 1995 survey.)

D. Summary of the Recommended Frame Development Procedures

To develop the F.W. Dodge sample frame, we propose to obtain two lists from F.W. Dodge, the first will be a census of large projects (250,000 square feet or more) and the second will be a sample of smaller projects (10,000 to 250,000 square feet). The large building frame will include projects with F.W. Dodge dates of 1995 through 1998. The moderate-sized building frame will

include projects with F.W. dodge dates 4/1/95 through 12/31/98. Both frames will be restricted geographically to the 129 CBECS PSUs.

The final F.W. Dodge frame will exclude building renovations and will exclude building additions that did not more than double the size of the building. The F.W. Dodge large building sample frame excludes buildings that were completed in 1995 if groundbreaking was prior to 1/1/95 and the moderate-sized building sample frame excludes buildings that were completed in 1995 if groundbreaking was prior to 4/1/95. Large buildings in the longitudinal frame that were completed in 1995 will be checked for inclusion on the 1999 F.W. Dodge frame. If a longitudinal frame building appears on the 1999 F.W. Dodge list, it will be assigned a PSU weight and will be deleted from the F.W. Dodge frame. There will be no special treatment for moderate-sized buildings in the longitudinal frame whose completion date was in 1995. It is assumed that the two frames do not overlap.

IV. PRELIMINARY SAMPLING RATES FOR THE F.W. DODGE SAMPLE

Three sets of sampling rates are required for the F.W. Dodge sample frame development and sample selection procedures: the F.W. Dodge list selection rate, the facility selection rate, and the building selection rate. We can use our experience with previous Dodge samples to estimate the appropriate selection rates for each stage of the sample. However, to some extent, the actual number of buildings obtained from the F.W. Dodge list will determine the final sampling rates.

A. Expected Yield of Target Rates for 1995

As a starting point, it is useful to examine the expected number of completes that would be obtained if the target rates from 1995 were applied to the 1999 F.W. Dodge List. This will give us a good idea of whether the sampling rates required to obtain 500 completes is likely to be higher than or lower than the sampling rates for 1995.

First, we consider the conceptual frame (i.e., buildings from 10,000 square feet to 250,000 square feet). Table 1 presents information on the expected sample yield for these buildings, assuming

that the same number of buildings were constructed each year from 1995 through 1998, as were constructed for the average of the years from 1990 through 1994. The first column of the table lists the size class of interest. The second column of the table shows the estimated number of buildings by size class built each year (based on the 1995 CBECS report). The third column shows the number that would be projected for 1995, 1996, 1997, and 1998. The fourth column shows the target sampling rates for 1995. The fifth column shows the expected sample yield using the targeted sampling rates and assuming an 80% response rate. Using the 1995 target rates, a total of 315 completed interviews would be expected from the conceptual frame (i.e. buildings from facilities between 10,000 and 250,000 square feet).

Table 1: Conceptual Frame Sample Yield Using 1995 Target Rates

Size Class	Actual Buildings per Year (90 to 94)	Expected 1995/1998 Buildings	1995 Target Rate	Final Sample Yield
10,000 - < 25,000	5,800	23,200	0.0025	46
25,000 - < 50,000	4,000	16,000	0.0075	96
50,000 - < 100,000	2,400	9,600	0.01	77
100,000 - < 250,000	800	3,200	0.0375	96

Next, we review the F.W. Dodge large building frame. In the 1995 CBECS, the three F.W. Dodge files, covering 1986 through 1994 (nine years) yielded 249 sample buildings, 151 of which were eligible for interview and 137 of which had a completed interview. That is an average of 15.2 building interviews per year. After applying a response rate reduction factor, we estimate that each year covered by the F.W. Dodge sample would yield 13.5 building interviews. Based on these statistics, we can estimate that the F.W. Dodge large building sample would yield 54 building interviews.

Together, the two sample frames can be projected to yield 369 interviews. This implies that the 1995 target sampling rates are too low to achieve the desired number of completes (i.e., 500

completed interviews). Sampling rates would have to be increased by approximately 35% to achieve the desired number of completes.

B. Setting Sampling Rates for Facilities in the F.W. Dodge List

For this preliminary analysis, we will assume that the appropriate dividing line for the F.W. Dodge large building frame (census) and the F.W. Dodge small and moderate building frame (sample) is 250,000 square feet. The large building frame is a census. Therefore the proposed sampling rate is 1.0. The small and moderate building frame is a sample. The sampling rate from the list must be at least as high as the maximum projected sampling rate for buildings in size class 5 (100,000 to 250,000 square feet).

In the previous section, we determined that using the target sample rates for 1995 would be inadequate to yield 500 cases from the F.W. Dodge sample. Assuming that we want to maintain the same relative structure of the sample, this implies that the sampling rate for size class 5 must be increased from 0.0375 to at least 0.05, to yield 500 cases.

F.W. Dodge is furnishing us with a complete file of building projects. We will use the information in this file to select a sample of facilities. We will then submit the F.W. Dodge project ID numbers to F.W. Dodge and obtain complete information for those construction projects. Our current contract allows us to select 3,500 reports. This should be adequate to allow us to sample enough projects in each size group to exceed our targeted sampling rates.

C. Setting Sampling Rates for Screening Facilities

In setting our initial sampling rate for screening facilities, we will choose aggressive targets, but will also select supplemental sample replicates. We will field the minimum number of facilities that we expect will be required to obtain the required number of buildings for the sample. At the same time, we will have additional, pre-selected sample replicates available for use if the selected facilities do not perform in the way that we expect.

D. Setting Sampling Rates for Screened Buildings

The final sampling rates for screened buildings will maintain the same relative structure used in the 1995 sample. The 1995 sampling rates by size class are shown in Table 2. The final column presents the preliminary sampling rates for 1999. The actual sampling rates used will be somewhat different, since the actual number of buildings constructed during the period 1995 to 1998 will be different from our projections. However, the same relative rates will apply.

Table 2: Target sampling rates by size class for 1995 sample

Size Class	1995 Target Sampling Rates	Preliminary 1999 Target Sampling Rates
10,000 - < 25,000	0.0025	0.0034
25,000 - < 50,000	0.0075	0.0101
50,000 - < 100,000	0.01	0.0135
100,000 - < 250,000	0.0375	0.0506
250,000 - < 500,000	0.075	0.1012
500,000 - < 1,000,000	0.15	0.2025
1,000,000 - < 4,000,000	0.25	0.3375
4,000,000 +	1.0	1.0000

The targeted number of sample cases for the F.W. Dodge sample is 1,042 cases. Of those, our 1995 CBECS experience shows that about 625 (60%) will be in-scope and eligible for the interview. If a response rate of 80% is achieved, the F.W. Dodge sample frame will deliver 500 completed interviews.

V. SELECTING SAMPLES FOR THE CBECS PRETESTS

The plan for the 1999 CBECS specifies two pretests, each consisting of 100 building interviews. In the pretests, EIA wishes to test both the reinterview procedures (i.e., fielding longitudinal sample cases) and the new sample building identification and interviewing procedures (i.e., F.W.

Dodge sample cases). To accommodate these tests, we will select a sample of longitudinal frame cases and a sample of F.W. Dodge frame cases.

A. Selecting Pretest Samples of Longitudinal Sample Cases

There were 6,590 sample cases eligible and in-scope for the 1995 CBECS. Adding the 21 cases where the building was under construction, we find that the available sample for the 1999 CBECS frame is 6,611 cases. Our preliminary targeted sample size is 5,921; leaving 690 cases available to furnish sample for the pretests.

Two special considerations affect the selection of the pretest samples.

First, to the extent possible, we would like to minimize the overlap between the pretest sample cases and the sample cases selected for the 1999 survey. If we implement the sampling procedure recommended in Section II, certain groups of cases will be subsampled for the 1999 CBECS with certainty. If it will not reduce the effectiveness of the pretest, groups of cases that are expected to have a subsampling rate of 1 will be excluded from the pretest sample.

Second, the goal of the pretests is to understand the effectiveness of the proposed 1999 CBECS procedures. Therefore, it is important to select pretest cases in a way that serves that purpose. We expect that there are a number of building-specific factors that will influence the effectiveness of the proposed procedures. The following are preliminary groups into which we propose to categorize buildings to help control selection for the pretests.

- *Size Groups* – Large, moderate, and small
- *Ownership Groups* – Owner occupied, leased
- *Building Use Groups* – (Categories to be determined)
- *Facility Groups* – Part of a facility, free-standing building
- *Energy Bill Groups* – Single energy bill, energy bill for each tenant

This is not intended to be a formal stratification mechanism. Rather, the goal would be to ensure that at least a few of the pretest cases fall into each cell. This can be achieved by picking a simple random sample of the available buildings and then iterating the sample selection to adjust the number of cases in each group.

We recommend selecting 100 cases from the longitudinal frame for each of the pretest samples. Of these, 65 would be initially fielded and the remainder would be held as a reserve sample. With a 95% eligibility rate and an 80% response rate, the 65 sample buildings should yield about 50 interviews. If the eligibility rate or the response rate is lower than anticipated, the reserve sample would be used to generate the additional interviews.

B. Selecting Pretest Samples of F.W. Dodge Sample Cases

The F.W. Dodge sample presents different challenges from the longitudinal sample. The most important difference is that, with the F.W. Dodge Sample, we must use a telephone interview to define the building in questions. For the longitudinal sample, on the other hand, an interviewer directly observed the building in 1995. One of the goals of the pretest is to ascertain the effectiveness of the set of procedures used to define a building.

Given the potential problems in defining a building, we recommend using a field verification procedure for the first pretest. The first pretest sample will be geographically clustered in four CBECS PSUs. We will send staff to conduct onsite verification of the information obtained for a building during the screening and interviewing process in two of the PSUs. By comparing the information obtained by telephone to the information obtained at the building location, we may be able to improve the building identification procedures. If significant problems are detected in the first two PSUs, we have the option of conducting onsite verification in the remaining two PSUs. If significant changes are made in procedures in an attempt to improve building identification, it may be appropriate to repeat the verification process for the second pilot test.

As noted in Section IV, we expect that 60% of F.W. Dodge buildings will be determined to be eligible and in-scope for the interview. Assuming an 80% response rate is achieved, the pretest

samples will each require 104 sample cases. We propose to select an initial sample of 150 cases. This will furnish a reserve sample to be used if the eligibility rate or the response rate falls below expectations.

VI. SUMMARY AND RECOMMENDATIONS

The 1999 CBECS will target about 5,000 completed interviews. About 4,500 interviews are to be obtained by reselecting cases that were selected for the 1995 CBECS. About 500 interviews are to be obtained by developing a sample frame from the F.W. Dodge list of new construction for the years of 1995, 1996, 1997, and 1998.

We recommend that the longitudinal sample frame for the 1999 CBECS include all eligible buildings from the 1995 survey, including both responding and nonresponding buildings. In addition, we recommend including 21 cases from the 1995 survey that were "out-of-scope" because the building was under construction. Preliminary estimates indicate that about 90% of the 1995 CBECS sample cases will be reselected for the 1999 survey. We propose to use a sampling procedure that increases sampling rates for those cases where the *targeted* sampling rate for 1995 was not realized. In addition, we will designate ten influential buildings from the 1995 CBECS as certainty selections.

We recommend that the F.W. Dodge sample frame cover all projects of at least 10,000 square feet that were constructed in 1995, 1996, 1997, and 1998. We recommend developing a complete sample frame of projects of at least 250,000 square feet and a "conceptual" frame of projects from 10,000 to 250,000 square feet. The new large building frame will "dovetail" with the 1995 CBECS Dodge list and will include projects from 1/1/95 through 12/31/98. Where longitudinal frame buildings overlap with the new large building frame, the longitudinal case will be assigned a PSU weight and the building will be deleted from the new large building frame. The moderate-sized building frame will dovetail with the area frame and will include projects from 4/1/95 through 12/31/98. It is assumed that there is no overlap between the two

frames. The frames will be geographically restricted to the CBECS PSUs. Building additions will be eligible for interview only if they more than double the size of the original building.

For the F.W. Dodge sample, we recommend using the same "relative" sampling rates as were used for the 1995 CBECS. Our preliminary estimates are that the 1995 sample rates would have to be increased by about 35% to yield 500 completed interviews.

We recommend selecting 100 cases from the longitudinal frame for each of the two pretests. We recommend procedures to minimize the overlap between the pretest samples and the 1999 CBECS samples.

We recommend selecting 150 cases from the F.W. Dodge sample frame for each of the two pretests. For the first pretest, we recommend selecting the sample from four CBECS PSUs to permit on-site verification of building identification procedures. For the second pretest, we recommend selecting a geographically disperse sample from the CBECS PSUs.