United States
Department of
Agriculture

Food and
Nutrition
Service

Alexandria, VA

THE FOOD STAMP PROGRAM QUALITY CONTROL SAMPLING HANDBOOK

FNS HANDBOOK 311

(REPRINTED THROUGH CHANGE NO. 2)

March 1990

Revisions to this Handbook *will be issued as revised* pages. They *will be issued with a transmittal sheet indicating Change I, Change* 2, etc. *Form FNS-221,* Record *of Changes, is included in the front of this* Handbook *for recording changes. Consistent use of the Record of Changes will assure that the* Handbook *is current and up* to *date.*

The date and number of the change will appear at the bottom left of the revised page. Asterisks in the margin will indicate lines that have been revised.

DISTRIBUTION:	RESPONSIBLE FOR PREPARATION AND MAINTENANCE:
F3,W (Special Distr.)	PAD-200

FORM FNS-378 (10-79)

UNITED STATES DEPARTMENT OF AGRICULTURE
FOOD AND NUTRITION SERVICE
ALEXANDRIA, VA 22302

CHANGE TRANSMITTAL	
CHANGE NUMBER: 2	DIRECTIVE IDENTIFICATION AND NUMBER: FNS Handbook 311
DIRECTIVE TITLE: The Food Stamp Program Quality Control Sampling Handbook.	

This Change transmits revised pages to correct typographical errors on pages 3-38, 3-40, and M-1.

PAGE CONTROL CHART

Remove Pages	Dated	Insert Pages	Dated
3-37 thru 3-40	Undated	3-37	Undated
		3-38	10-18-90
		3-39	Undated
		3-40	10-18-90
Exhibit M, M-1 and M-2	Undated	M-1	10-18-90
		M-2	Undated

PHYLLIS R. GAULT
Acting Deputy Administrator
for Food Stamp Program

DISTRIBUTION:	MANUAL MAINTENANCE INSTRUCTIONS:	OPI:	Change 2 10-18-90
F3,W (Special distr.)	Retain and file Change Transmittal in front of directive until further notice. Remove and insert page(s) if applicable.	PAD-200	

Form FNS-389 (9-88) Previous editions obsolete

UNITED STATES DEPARTMENT OF AGRICULTURE

FOOD AND NUTRITION SERVICE

ALEXANDRIA, VA 22302

CHANGE TRANSMITTAL	
CHANGE NUMBER: 2	DIRECTIVE IDENTIFICATION AND NUMBER: FNS Handbook 311
DIRECTIVE TITLE: The Food Stamp Program Quality Control Sampling Handbook.	

This Change transmits revised pages to correct minor text errors on pages 3-7 and E-10. The correction on pages 3-7 corrects an entry error in Figure 3-4 to conform with 7 CFR 275.11(b) (3). The correction on page E-10 corrects typographical errors in the first equation on the page.

PAGE CONTROL CHART

Remove Pages	Dated	Insert Pages	Dated
3-7 and 3-8	Undated	3-7 3- 8	8-17-90 Undated
Exhibit E, E-9 and E-10	Undated	E-9 E-10	Undated 8-17-90

Thomas O'Connor

for BONNY O'NEIL
Acting Deputy Administrator
for Food Stamp Program

DISTRIBUTION:	MANUAL MAINTENANCE INSTRUCTIONS:	OPI:	Change 1 8-17-90
F3,W (Special distr.)	Retain and file Change Transmittal in front of directive until further notice. Remove and insert page(s) if applicable.	PAD-200	

Form FNS-389 (9-88) Previous editions obsolete

RECORD OF CHANGES *FNS Instr. 122-1*	USDA-FNS	NAME AND/OR NUMBER OF PUBLICATION FNS HANDBOOK 311

NOTE: As changes are transmitted, record the change number, the date of change, and the date filed.

CHANGE NO	DATE OF CHANGE	DATE FILED	CHANGE NO.	DATE OF CHANGE	DATE FILED

FORM FNS-221 (10-79) *Previous editions may be used until supply is exhausted.*

UNITED STATES DEPARTMENT OF AGRICULTURE
Food and Nutrition Service
3101 Park Center Drive
Alexandria, VA 22302

FNS NOTICE 90-23

ACTION BY: State Agencies

INFORMATION FOR: Regional Offices

FNS Handbook 311, The Food Stamp Program
Quality control Sampling Handbook

This Notice transmits FNS Handbook 311, The Food Stamp Program Quality Control Sampling Handbook. This Handbook replaces FNS Handbook 311 issued in August 1979. Many of the Handbook's revisions are policy changes that have been promulgated by FNS in regulations published since August 1979. The most important of these regulations was the one published February 17, 1984. It provided for the implementation of an annual sampling period, a reduction in the minimum sample size requirements, and a requirement that State agencies agree to accept the level of reliability of error rates resulting from the sample sizes which they select. This regulation also established deadlines for submitting State sampling plans.

In addition to reiterating policy, FNS Handbook 311 also contains background information on sampling concepts and provides step-by-step instructions for sampling activities. The revised Handbook provides additional details in these areas, especially in conjunction with integrated sampling and the operational aspects of stratified sample designs. The revised Handbook also updates the instructions for completing the February 1990 revision of Form FNS-247, Statistical Summary of Sample Disposition, and the November 1989 revision of Form FNS-248, Status of Sample Selection and Completion. The Attachment to this Notice provides a detailed outline of the revisions to FNS Handbook 311.

Remove from Manual and destroy FNS Handbook 311 issued in August 1979. Insert this Handbook.

Bonny O'Neil

BONNY O'NEIL
Acting Deputy Administrator
for Food Stamp Program

Attachment

DISTRIBUTION:	EXPIRATION:	RESPONSIBLE FOR PREPARATION AND MAINTENANCE:	3-15-90
F3,W (Special distr.)	Destroy Notice when action is completed.	PAD-200	

Form FNS-619 (4-86)

CHANGES IN FNS HANDBOOK 311

1. Changes in style

    1.1. Assignment of an identification number to each equation

    1.2. Placement of highly-mathematical material in Exhibits

    1.3. Increased use of personal pronouns: "you" for the State agency; "we" for FNS

2. Changes in content

    2.1. Additions to discussion of sampling concepts:

        2.1.1. Sampling-frame problems

        2.1.2. Sampling from overlapping populations

    2.2 Additions/changes to material on sampling activities:

        2.2.1. Additional material on prediction of caseloads

        2.2.2. Distinctions between:

            2.2.2.1. Minimum sample size based on estimated caseload

            2.2.2.2. End-of-year minimum sample size

            2.2.2.3. Desired sample size

            2.2.2.4. Required sample size

        2.2.3. Minimum sample sizes vis-a-vis demonstration and SSA-processed cases

        2.2.4. Additional details on sampling plans:

            2.2.4.1. Requirement to retain supporting documents and records of sampling activities

            2.2.4.2. Writing guideline

            2.2.4.3. Contents of integrated plans

2.2.4.4. Submission of plans (definition of major/minor changes)

2.2.5. Additional details on frame construction

2.2.6. Description of an alternative method (replicated sampling) for systematic sampling with a decimal interval

2.2.7. Additional details on simple random sampling

2.2.8. Update to forms and instructions for Form FNS-247, Statistical Summary of Sample Disposition, and Form FNS-248, Status of Sample Selection and Completion

2.2.9. New material on determination of actual out-of-scope rate and actual subject-to-review caseloads

2.2.10. New material on end-of-year reconciliation

2.2.11. Additional detail on sampling corrections

2.2.12. New material on FNS reviews of sampling procedures

2.3. New material in Exhibits:

2.3.1. Bibliography

2.3.2. Summaries of integrated sample designs

2.3.3. Formulas for dollar error rates

2.3.4. Comparing precisions of stratified and unstratified sample designs

2.3.5. Estimation of stratum weights for stratified samples

2.3.6. Reliability waiver statement

2.3.7. Computer generation of random numbers

2.3.8. New table of random digits

United States
Department of
Agriculture

Food and
Nutrition
Service

Alexandria, VA

THE FOOD STAMP PROGRAM QUALITY CONTROL SAMPLING HANDBOOK

FNS HANDBOOK 311

March 1990

Revisions to this Handbook *will be issued as revised* pages. They *will be issued with a transmittal sheet indicating Change I, Change* 2, etc. *Form FNS-221,* Record *of Changes, is included in the front of this* Handbook *for recording changes. Consistent use of the Record of Changes will assure that the* Handbook *is current and up* to *date.*

The date and number of the change will appear at the bottom left of the revised page. Asterisks in the margin will indicate lines that have been revised.

DISTRIBUTION:	RESPONSIBLE FOR PREPARATION AND MAINTENANCE:
F3,W (Special Distribution)	PAD-200

FORM FNS-378 (10-79)

USDA-FNS

RECORD OF CHANGES

FNS Instr. 122-1

NAME AND/OR NUMBER OF PUBLICATION

FNS HANDBOOK 311

NOTE: As changes are transmitted, record the change number, the date of change, and the date filed.

CHANGE NO	DATE OF CHANGE	DATE FILED	CHANGE NO.	DATE OF CHANGE	DATE FILED

FORM FNS-221 (10-79) *Previous editions may be used until supply is exhausted.*

TABLE OF CONTENTS

PAGE

CHAPTER I. INTRODUCTION

CHAPTER II. SAMPLING CONCEPTS

CHAPTER III. SAMPLING ACTIVITIES

PAGE

FIGURES

EXHIBITS

CHAPTER I

INTRODUCTION

1000 GENERAL. The Food Stamp quality control (QC) system produces summary information on Food Stamp certification errors. This information is of interest to State and federal governments as well as to the general public. The Food Stamp QC system consists of the following tasks:

A Sample frame construction,

B Sample selection,

C QC data collection, and

D Error rate estimation.

Three Food and Nutrition Service (FNS) handbooks describe QC activities. FNS Handbook 310 describes QC-data collection. This handbook, FNS Handbook 311, describes sample-frame construction, sample selection, and the determination of State-calculated error rates. FNS Handbook 315 describes the determination of Federal regressed error rates.

The intended audience for this handbook is State agency employees involved with the QC sampling of Food Stamp cases. Consequently, the word "you" in this handbook refers to the State agency. The word "we" refers to FNS, and the word "year" refers to the federal fiscal year: October 1 through September 30.

1100 HANDBOOK ORGANIZATION. This handbook consists of three major sections: chapter II, chapter III, and the exhibits. Chapter II describes sampling concepts; it defines terms and introduces general principles of sampling. Chapter III describes sampling activities; it presents a step-by-step checklist for a State's QC sampling of Food Stamp cases. The exhibits provide supplementary information on various topics. For example, Exhibit A, attached, lists alphabetically the definitions appearing in other sections of the handbook.

1200 SOURCE MATERIALS. In preparing this handbook, we used two types of source materials: federal regulations and textbooks on sampling. The bibliography in the attached Appendix cites the Food Stamp regulations that apply to QC sampling and lists some sampling textbooks.

CHAPTER II

SAMPLING CONCEPTS

2000 GENERAL. This chapter introduces terminology and concepts associated with probability sampling procedures.

2100 UNIVERSES AND FRAMES.

2110 TERMINOLOGY. To discuss sampling concepts, we must first introduce some terminology. "Elements" are the elementary units for which you seek information. The element for the Food Stamp active error rate is an active case[1]; that is, a household participating in the Food Stamp program for a specified month. The element for the Food Stamp negative-error rate is a negative action[2], which is a household that is denied or whose benefits are terminated.

The "target population" is the set of all elements of interest. The target populations for the Food Stamp program are all active and negative cases for a fiscal year, except those excluded as not subject to review as defined in 7 CFR 275.11(f). Another term for target population is "universe".

To permit sampling, a population must be subdivided into a finite number of distinct and identifiable units called "sampling units", which are each composed of an integer (possibly zero) number of elements. A "sampling frame" is the physical representation of the target population. A sampling frame is usually a list of sampling units. If there is no such list, a sampling frame is some equivalent procedure for identifying sampling units. An example of the latter is random digit sampling (see 3222 C).

2120 UNIVERSES AND FRAMES IN FOOD STAMP QC. In Food Stamp QC, sampling frames are nearly always lists of active and negative cases. The sources of these lists vary from State to State. Some examples for the active frame are lists of participants, issuances, or certified households. (See 3210.)

The universe may be represented by more than one list. For example, the frame for active cases can be comprised of 20 different lists representing daily issuances on each working day of the month. The negative frame is often composed of one list of denials and another list of terminations.

---

1 See FNS Handbook 310, section 130.

2 The QC regulations and Handbook 310 (section 1320) use the phrase "negative case" instead of "negative action". "Negative case" is a derived term created to achieve parallel regulatory language about the sampling activities associated with active and negative error rates.

2130 SAMPLING-FRAME PROBLEMS. Ideally, there is a one-to-one correspondence between elements of the target population and sampling units in the frame[3]. If such a relationship does not exist, there are problems with the sampling frame. The four most common problems with sampling frames are the following:

A Omission of one or more population elements from the sampling frame. This causes some portion of the population to have no chance of selection or review (i.e., probability of selection equal zero). This can lead to biased error-rate estimators (see 2300).

B Inclusion in the sampling frame of units that are not part of the target population. These units are said to be "out of scope". In Food Stamp samples such cases are to be coded as Not Subject To Review.

C Duplication of population elements in the sampling frame. This is a result of some sampling units appearing more than once. This often occurs when several lists are used to represent the target population but can also occur within a single list. This also can lead to biased error-rate estimators.

D Allowing sampling units to consist of more than one population element[4]. This is often referred to as "clustered sampling units". If a sampling frame contains clustered sampling units, then either (1) a sample selected from the sampling frame must be treated as a cluster sample or (2) clustered sampling units must be repeated in the sampling frame such that each occurrence corresponds to a unique population element.

2140 FRAMES FOR OVERLAPPING POPULATIONS. A "domain" is a subset of a population. Two or more populations "overlap" if they contain a set of common elements, referred to as the "overlap domain". The integration of Food Stamp, AFDC, and Medicaid quality control reviews is possible because these three populations overlap.

You can develop sampling frames for overlapping populations in a variety of ways. One way of characterizing such frames is by the following descriptions of the interrelationships between sampling frames:

A "Totally integrated frame"--one sampling frame containing all the sampling units.

B "Overlapping sampling frames"--two or more sampling frames containing some common sampling units.

---

3 The sampling frames for cluster sampling (see Exhibit B, attached) are an exception to this statement.

4 This is not a sampling-frame problem in cluster sampling.

C "Disjoint sampling frames"—two or more sampling frames containing no common sampling units.

Regardless of the type of frame(s) available to represent overlapping populations, you should still be aware of the possible sampling frame problems outlined in 2130. Every population element must be contained (preferably only once) in the frame, or combination of frames, representing each target population. For example, if a household receives both Food Stamp and AFDC benefits, it must be included in both the Food Stamp frame(s) and the AFDC frame(s).

The sampling frames for overlapping populations may have "control data" associated with each sampling unit. This identifies the domain to which each sampling unit belongs and simplifies integrated sampling (see 2250). If this type of information is available, additional problems can arise if the control data is inaccurate. This could lead to the assignment of population elements to an inappropriate frame and/or stratum.

2200 SAMPLE DESIGNS.

2210 TERMINOLOGY. "Sampling" is the selection of a subset of units, called a "sample", to represent a larger set of units. A "sampled unit" is a sampling unit that has been selected to be in a specific sample. Sampling of a sample is called "subsampling". The "size" of a sample is the number of sampled units that it contains. The purpose of sampling is to reduce the amount of data needed to develop quantitative information.

A "sampling procedure" is a sequence of operations that produces a sample. "Sampling parameters" are numerical quantities that characterize a sampling procedure. The size of a sample is an example of a sampling parameter.

A "sample design" describes one or more populations and the associated sampling procedure(s). A sample design need not specify the values for the sampling parameters but may instead describe how these values are to be determined. This is in contrast to a "sampling plan" (see 3130), which must specify the values for sampling parameters.

In "element sampling" each sampling unit contains no more than one element. This is in contrast to "cluster sampling" (see Exhibit B), in which the sampling units are clusters of several (or many) elements. An element-sampling scheme in which every sampling unit has equal probability of being sample selected is called an equal-probability selection method, or "epsem". In an epsem design the size of the sampling frame is usually denoted by N, or N(frame), and the size of the sample is usually denoted by n. The fraction $f = n/N$ is called the "sampling fraction" or "sampling rate". Two examples of epsem's are (1) simple-random sampling of elements and (2) systematic sampling of elements.

2220 SIMPLE RANDOM SAMPLING. In "simple random sampling" any one of the possible subsets of n distinct sampling units chosen from the sampling frame of N elements is equally likely to be a sample of size n. You select a simple random sample by selecting random numbers from a table or by generating them with a computer program. To select a simple random sample, you proceed as follows:

A Obtain a random number k between 1 and N (inclusive).

B If this is the first time that random number k has been obtained, add the k^{th} sampling unit to the sample.

C Repeat the above steps until the sample contains n sampled units.

Because you check each random number for previous use in Step B, this is called "sampling without replacement"; that is, after you use a random number, you do not "replace" it in the pool of available random numbers.

2230 SYSTEMATIC SAMPLING. An alternative epsem procedure is "systematic sampling". If you sample manually, it is usually easier to perform than simple random sampling. The key sampling parameter is the "sampling interval", which is the reciprocal of the sampling rate. If the sampling interval, denoted q, is an integer, you select a systematic sample by proceeding as follows:

A Obtain a random number k between 1 and q (inclusive) and add the k^{th} sampling unit to the sample.

B Add to the sample the sampling units k+q, k+2q, k+3q, etc. up to the largest possible such sampling unit from the sampling frame.

When the sampling interval is not an integer, you select a systematic sample by using one of the methods described in 3222 A.

Systematic sampling differs from simple random sampling in that the probabilities of different sets of elements being included in the sample are not all equal. For example, if the sampling interval is 482 the probability that elements 1 and 2 are both in the sample is zero, whereas the probability that elements 1 and 483 are both in the sample is 1/482, since if element 1 is in the sample then element 483 is bound to be as well. On the other hand, in a simple random sample the probability for any two distinct elements being in the sample is the same for every pair of elements.

2240 STRATIFIED SAMPLING. An element-sampling scheme that may or may not be an epsem procedure is stratified sampling. "Stratification" is the partitioning of a sampling frame into disjoint and exhaustive sets of sampling units. The resulting sets of sampling units are called "strata" (singular: "stratum"). In "stratified sampling" you select a separate, independent sample of a specified size from each stratum.

In stratified sampling, the size of the h^{th} stratum is usually denoted N_h, or N_h(frame), and the size of the sample selected in the h^{th} stratum is usually denoted n_h. The stratum sampling fraction is thus $f_h = n_h / N_h$. The sum of the N_h's is usually denoted by N, and the sum of the n_h's by n. The specification of the stratum sample sizes—that is, the breakdown of n into the n_h's—is called "allocation". In "proportional allocation", all of the stratum sampling fractions are equal. This is so called because the n_h's are proportional to the N_h's. The absence of all the stratum sampling fractions being equal is called "disproportionate allocation". Some of the reasons for using a stratified sampling are to increase precision (see 2300), to permit integrated sampling of overlapping populations (see 2250), to permit a change in sampling rates when sampling on more than one occasion (see 3414), or to permit estimation of population characteristics for each stratum.

2250 INTEGRATED SAMPLING OF OVERLAPPING POPULATIONS. In an "integrated sample design" some (or all) of the sampled units receive combined data collection for more than one population. For example, if the Food Stamps and AFDC QC samples both have sample sizes of 1200, then an integrated design might consist of 500 Food-Stamp-only reviews, 500 AFDC-only reviews, and 700 joint reviews. The total number of reviews for this integrated design is then 500+500+700 = 1700, which is less than the 1200+1200 = 2400 reviews for non-integrated design. Hence, integrated sampling reduces data collection costs. There are several sampling methods, used separately or in combination, for producing integrated sample designs. These include stratified sampling, cross-classification, and sample replacement.

One method for producing an integrated sample design is the use of stratified sampling. The strata used are domains or sets of domains. You then subject to joint data collection some (or all) of the sampled units occurring in the overlap domains. This approach requires either disjoint sampling frames or the existence of control data in a combined sampling frame.

A second method for producing an integrated sample design is the use of "cross-classification", which is the determination of a sampled unit's domain membership after sample selection. Cross-classification partitions the sample into "subclasses" that correspond to the population domains. Like in stratified sampling, you then subject to joint data collection some (or all) of the sampled units that you have assigned to the overlap subclass(es). You should use cross-classification instead of stratified sampling when your sampling frames are overlapping and do not contain control data.

A third method for producing an integrated sample design is the use of "sample replacement". If you have two or more overlapping populations plus the corresponding sets of sampled units contained in the overlap, sample replacement designates one set of sampled units as subject to data collection for all or various combinations of the populations and excludes from all data collection the remaining set(s) of sampled units. The usual sequence of operations is the following:

(1) selection of samples, one per population, from overlapping sampling frames; (2) cross-classification of each set of sampled units; (3) in some cases, subsampling into replacement subset(s) and non-replacement subset(s); and (4) sample replacement.

The Integrated Quality Control System sampling handbook (published December 1979) describes eight integrated sample designs. Exhibit C, attached, summarizes these designs.

2300 PROPERTIES OF SAMPLE-BASED ESTIMATORS.

2310 TERMINOLOGY. As previously mentioned in 2210, the purpose of sampling is to reduce the amount of data needed to develop quantitative information. Thus, sampling reduces the cost of information. If, hypothetically, you had an inexhaustible data collection budget, you could collect data for all population elements and then use the resulting data to calculate "population parameters", such as population means and totals. This, however, is usually prohibitively expensive. Consequently, you use sampling and then use the resulting data to calculate estimates that approximate the population parameters of interest. The term "estimate" refers to a specific value that results from a computation involving sample data. The term "estimator" refers to the computational procedure used for obtaining the estimate. Exhibit D, attached, describes estimators for State-estimated error rates.

An estimate's "expected value" is the average value of the estimates resulting from all possible samples that can be selected in accordance with your sampling plan. If an estimate's expected value differs from the population parameter being estimated, the estimate is said to be "biased". The amount of this difference is called the estimate's "bias". Unbiased estimates are preferred because then the resulting estimates equal "on the average" the population parameter being estimated.

The term "accuracy" is an opposite-sense description of bias. If an estimate's bias is large, the estimate is said to have low accuracy. If on the other hand, an estimate's bias is small, the estimator is said to have high accuracy. The term "validity" is synonymous with accuracy.

Even if an estimate is unbiased, any one estimate will nearly always be different from the population parameter being estimated. Over all possible samples, some estimates will be larger than the population parameter whereas other estimates will be smaller. A measure of this variability is the estimate's "variance". This is the average over all possible samples of the squared difference between the resulting estimate and the expected value of the estimator. The square root of the estimate's variance is called the "standard error" of the estimator.

The term "precision" is an opposite-sense description of variance. If an estimate's variance is large, the estimate is said to have low precision. If, on the other hand, an estimate's variance is small, the estimate is said to have high precision. This latter situation—high precision, or low variance—is preferred because then over repeated samples the variability of resulting estimates is small. One estimate is said to have "better precision" than another if it has higher precision—that is, smaller variance—than the other estimate. Exhibit E, attached, discusses procedures for determining if one estimate has equal or better precision than another estimate. The term "reliability" is synonymous with precision.

"Estimation error" is the difference between an estimate and the population parameter being estimated. Estimation error can be decomposed into sampling error and possible non-sampling errors. "Sampling error" occurs when you collect data for only a sample and not the entire population. "Non-sampling errors" can be attributed to many sources. These include definitional difficulties, differences in interpretation, inability or unwillingness of respondents to provide correct information, mistakes in data recording, and other errors in collection, response, processing, coverage, and estimation for missing data.

Exhibit F, attached, contains an illustration of the properties of sample-based estimators defined above.

2320 EFFECTS OF SAMPLE DESIGN. The properties of sample-based estimators are affected by features of both sample design and data collection procedures. The effects of sample design arise from characteristics of the population(s), sampling frame(s), and sampling procedure(s). For example, sampling-frame problems (see 2130) can be a source of estimator bias. In addition, several sample-design features can increase estimator precision. These include stratification, increases in sample size, and sample allocations that increase sample sizes in heterogeneous strata while decreasing sample sizes in homogeneous strata.

CHAPTER III

SAMPLING ACTIVITIES

3000 GENERAL. This part of the handbook (3000 through 3400) describes sampling activities. We have organized the described sampling activities in time sequence: 3100 describes planning activities; 3200, monthly activities; 3300, end-of-year activities; and 3400, other activities performed during the year.

3100 PLANNING ACTIVITIES. Planning activities include: the prediction of caseloads, determination of sampling parameters, and the preparation of the sampling plan.

3110 PREDICTION OF REVIEWABLE CASELOAD SIZES.

3111 GENERAL. A necessary planning activity is the prediction of the approximate sizes of the active and negative target populations. In other words, you must predict the active and negative caseloads subject to review for quality control. These predictions, in turn, determine a number of sampling parameters, such as minimum sample size (see 3121) and the sampling interval in systematic sampling (see 3122).

Though the target populations consist of cases for the entire year, the subject-to-review caseloads you must predict are the average-monthly caseloads that are subject to review. There are two basic approaches to choose from for doing this: the "direct approach" or the "sampling-frame approach".

In the direct approach, you use historical subject-to-review caseload numbers to predict the average-monthly caseload that will be subject to review in the coming year. You obtain the needed historical numbers by applying one of the reviewable-caseload equations (see 3320) to your QC data for previous years. You then apply to these historical numbers one of the prediction method described in 3112 to obtain a prediction of the average-monthly caseload subject to review for the coming year.

In the sampling-frame approach, you first make two predictions (see 3112) for the coming year: a prediction of N_e', the estimated average-monthly frame size, and a prediction of d, the out-of-scope rate (see 2130(B)). You then use the following equation to calculate

$$N_e = \text{estimated average-monthly caseload subject to review}$$

$$= N_e'(1-d) \quad . \qquad (3\text{-}1)$$

The historical numbers used to predict N_e' are the annual averages of exact monthly frame sizes from previous years. The historical numbers used to predict d are the out-of-scope rates calculated from previous years' samples (see 3310).

You can use either approach. We generally prefer that you use the sampling-frame approach, however, because it "divides and conquers" your prediction problem by splitting it into two easier problems.

3112 PREDICTION METHODS. The following are some methods you can choose from to predict the average-monthly caseload subject to review, the average-monthly frame size, or the out-of-scope rate for the coming year:

A One of the easiest is to use the actual number for the current year as the predicted number for the coming year. This is appropriate if the number being predicted is expected to remain constant because of (1) the absence of observed trends in the number and (2) the absence of any known or anticipated changes effecting the number.

B A modification of the preceding method is to predict the coming year's number from two historical numbers: that for the current year and that for the immediately-previous year. In this method, the percentage increase or decrease between the previous and current years is applied to the current year number to yield a predicted number for the coming year.

For example, if (in the direct approach) the subject-to-review caseloads increased 11 percent between the previous and current years, then an 11 percent increase is applied to the average-monthly subject-to-review caseload subject to review for the current year to predict a subject-to-review caseload for the coming year.

This method is appropriate if the trend from the previous year to the current year is expected to continue with the same rate of increase or decrease for another year.

C You can also use three or more historical numbers to predict the number for the coming year.

For example, if (in the sampling-frame approach) we denote the current year as Year 3, and average-monthly frame size increased 15 percent from Year 1 to Year 2 but only increased 10 percent from Year 2 to Year 3, then a continuation of this trend would predict an increase of five percent for the change in average-monthly frame size from the current year to the coming year. Hence, applying a five percent increase to the average-monthly frame size for the current year would yield a predicted average-monthly frame size for the coming year.

This method is appropriate if there is a trend in the rate of increase or decrease, and this trend is expected to continue for another year.

D There are a number of prediction methods that use historical data in the form of individual monthly numbers instead of the previous years' average-monthly numbers. These methods include trend analysis of monthly data, regression modeling[5], and various time-series methods[6]. The trend analysis of monthly data is appropriate if there is a trend in individual monthly numbers that is expected to continue through all 12 months of the coming year. Regression modeling and time-series methods require the use of special computer programs. These two methods should not be used if you lack appropriate statistical training to thoroughly understand the underlying theory and assumptions. Prediction methods that use individual monthly data usually make predictions for individual months. If this is the case, you should predict individual numbers for each of the 12 months of the coming year and then average these together to yield the predicted average-monthly number.

E In some situations the use of a completely statistical procedure may not be appropriate. For example, if it is known that in the coming year a policy change effecting participation levels will occur, but this particular policy change has never occurred previously, then historical information will probably not be very useful in predicting the caseload for the coming year. In this kind of situation, your caseload prediction for the coming year will have to be based mostly on informed judgement.

Your choice of prediction method should be based on the applicability of the assumptions associated with each method, the desired degree of prediction accuracy (see 3114), and your statistical capabilities.

3113 PREDICTING REVIEWABLE CASELOADS FOR OVERLAP DOMAINS. Many integrated sample designs require that you predict reviewable caseloads for overlap domains prior to calculating certain sampling parameters. Like the caseload predictions for non-overlapping populations, there are basically two approaches to predicting caseloads subject to review in overlap domains.

In the sampling-frame approach, you first predict the number of overlap cases in the sampling frame. This is feasible if the current year's sampling frame contains control data, and thus you can obtain an exact count of the number of cases in the sampling frame contained in the overlap. Use this to predict the number of overlap cases contained in the coming year's sampling frame. Next, estimate from sample data the out-of-scope rate, d, for the overlap domain. This is based on the number of not-subject-to-review (NSR) cases in the sample data. Since you dispose a case as NSR if (1) it is out of scope or (2) has been

5 An example of using regression modeling to predict AFDC caseloads for Washington State is described by Plotnick and Lidman (1986) and for New Jersey by Barnow and Garasky (1988). (See the Appendix for complete citations.)

6 An evaluation of time-series methods to predict AFDC caseloads is described by Williams (1987). (See the Appendix for complete citation.)

deselected to correct for oversampling, the estimation of the out-of-scope rate must disregard cases that are NSR because of a deselection to correct for oversampling. Finally, you multiply the prediction of the number of overlap cases in the sampling frame by (1.0 - d) to yield a prediction of the number of overlap cases subject to review.

If your sampling frame lacks control data, you will not be able to use the above sampling-frame approach. Consequently, in the "sample-fraction approach" you first choose one of the populations and predict its caseload subject to review (See 3112.). For example, for the Food-Stamp/AFDC domain you could choose the AFDC population and first estimate the AFDC caseload subject to review. Next, estimate (from sample data) the proportion of cases in the chosen population that are also in the overlap domain. Finally, multiply the predicted caseload subject to review for the chosen population by the estimated overlap proportion to produce a prediction of the caseload subject to review for the overlap domain.

3114 EFFECTS OF INACCURATE PREDICTIONS. Your prediction of the reviewable caseload determines other sampling parameters: the minimum sample size (see 3121) and the sampling interval for systematic sampling (see 3122). Consequently, if you are sampling systematically (and do not correct for under- or oversampling), an inaccurate prediction of the caseload subject to review will result in either (1) the number of selected cases that are reviewable being less than your minimum sample size or (2) the number of selected cases that are reviewable being more than your minimum sample size. Which of these will occur depends on the following factors: (1) the magnitude and direction of the prediction error; (2) the beginning- and end-of-year minimum sample size; (3) the type of cases—active or negative; and (4) whether you opted for the reduced (active-case) sample size.

The table in figure 3-1 indicates the effect of an error in predicting the reviewable caseload when you (1) sample systematically, (2) calculate the sampling interval in the manner described in 3122[7], and (3) do not perform any corrections for under- or oversampling. A positive entry in the table indicates an expected excess number of reviewable cases, whereas a negative entry indicates an expected shortfall in the number of reviewable cases. The table in figure 3-1 contains two columns for the under-prediction of the reviewable caseload—one for less than 20 percent and the other for 20 percent or more. The reason for this is that your minimum sample size does not change if the under-prediction is less than 20 percent (see 3121.1)[8].

7 In particular, the discussion in 3114 assumes that the sampling interval is calculated (without rounding) by dividing the predicted average-monthly caseload by one twelfth of the particular minimum sample size corresponding to the predicted average-monthly caseload. Also, see 3121.1 B.

8 7 CFR 275.11b(3).

Figure 3-1

EFFECT OF INACCURATE CASELOAD PREDICTION ON THE NUMBER OF SELECTED CASES SUBJECT TO REVIEW IN THE ABSENCE OF ANY CORRECTIONS FOR UNDER- OR OVERSAMPLING

Multiply appropriate entry by number of percentage points of under- or over-prediction of the average-monthly caseload subject to review. A positive result indicates expected number of selected cases subject to review over the minimum sample size at the end of the year. A negative result indicates expected number of selected cases subject to review under the minimum sample size at the end of the year. For example, if (1) you are sampling negative cases, (2) both the beginning- and end-of-year minimum sample sizes are 800, and (3) you over-predict the subject-to-review caseload by nine percent, then the result is (9)x(8.00) = -72; in other words, (without a correction for undersampling) you can expect to be short 72 reviewable cases at the end of the year.

Type of Case	Reduced sample size	Minimum sample size*	Effect of over-prediction	Effect of under-prediction: less than 20 %	Effect of under-prediction: 20% or more
Active	Yes	1200	-12.0	+12.0	+12.0
Active	Yes	301 - 1199	-1.2	+np/100†	+1.2
Active	Yes	300	-3.0	+3.0	+3.0
Active	No	2400	-24.0	+24.0	+24.0
Active	No	301 - 2399	+1.2	+np/100†	-1.2
Active	No	300	-3.0	+3.0	+3.0
Negative		800	-8.0	+8.0	+8.0
Negative		151 - 799	-0.78	+np/100†	+0.78
Negative		150	-1.5	+1.5	+1.5

*Interval containing both (1) minimum sample size determined by the predicted caseload and (2) minimum sample size determined by the actual caseload. The table cannot be used if these two minimum sample sizes correspond to different table entries.

† np = minimum sample size determined by predicted average-monthly caseload subject to review.

3120 DETERMINATION OF SAMPLING PARAMETERS. Another necessary planning activity is the determination of the sampling parameters to be used in the coming year. This includes the determination of sample sizes and additional quantities that control the sampling process. For systematic sampling, the additional sampling parameters include the sampling interval and all random starts for the coming year.

3121 SAMPLE SIZES.

3121.1 MINIMUM SAMPLE SIZES. For systematic and simple-random sampling, FNS regulations specify "minimum sample sizes"—that is, the minimum number of cases to be reviewed during the year[9]. For active cases, there are two minimum-sample-size schedules: a reduced schedule and a non-reduced schedule. The reduced schedule applies if you submit a waiver statement in which you agree not to contest the final error rates on the basis of the precision resulting from the reduced sample size[10]. (See Exhibit G, attached, for an example of a waiver statement.) The following is the reduced sample-size schedule for active cases:

Figure 3-2

REDUCED MINIMUM SAMPLE SIZES FOR ACTIVE CASES

Average Monthly Reviewable Caseload (N)	Minimum Annual Sample Size (n)
60,000 or greater	1200
10,000 to 59,999	n = 300 + [0.018 (N - 10,000
less than 10,000	300

If you do not opt for a reduced active sample, the following schedule applies:

Figure 3-3

NON-REDUCED MINIMUM SAMPLE SIZES FOR ACTIVE CASES

Average Monthly Reviewable Caseload (N)	Minimum Annual Sample Size (n)
60,000 or greater	2400
10,000 to 59,999	n = 300 + [0.042 (N - 10,000)]
less than 10,000	300

9 7 CFR 275.11(b). For the discussion of minimum sample sizes for alternative sample designs see 3121.7 and Exhibit E.

10 7 CFR 275.11(a)(viii).

For negative actions, only the following schedule is available:

Figure 3-4

MINIMUM SAMPLE SIZES FOR NEGATIVE ACTIONS

Average Monthly Reviewable Caseload (N)	Minimum Annual Sample Size (n)
5,000 or greater	800
500 to 4,999	$n = 150 + [0.144\ (N - 500)]$
less than 10,000	150

In all three of the above figures, the average reviewable caseload (N) and the minimum sample size (n) are defined as follows:

A Definition of N. The reviewable caseload, N, is the average monthly size of the QC universe as described in 2120. It pertains only to households in the sampling frame that are subject to review. This is NOT the size of the frame but rather the frame size adjusted downward based upon the portion of the sampling frame that is out of scope. 3110 discusses how to predict reviewable caseloads.

B Determining n. At the beginning of the year, the minimum sample size will be based upon an estimate of the reviewable caseload, N_e. At the end of the year, the minimum sample size may change depending on the relationship between N_e and the actual reviewable caseload, N_a[11]. (See 3320 on how to calculate N_a at the end of the year.) Figure 3-5 explains how to determine the end-of-year minimum sample sizes for systematic and simple-random samples that are unstratified or proportionally stratified. (For alternative sample designs, see E330.) The final minimum sample size, n_f, may vary in either direction from the originally calculated figure, n_e, when the actual caseload is either less than the estimated caseload (table entry a) or more than twenty percent greater than the estimate (table entry d). If the ratio of actual to estimated caseload is between 1 and 1.2 (table entries b and c), the minimum sample size is unchanged.

Part A of Figure 3-6 shows an example of the calculation of n_e; Part C, an example calculation of n_f.

11 7 CFR 275.11(b)(3).

Figure 3-5

DETERMINATION OF THE END-OF-YEAR MINIMUM SAMPLE SIZE FOR SYSTEMATIC AND SIMPLE-RANDOM SAMPLES THAT ARE UNSTRATIFIED OR PROPORTIONALLY STRATIFIED

entry	situation	result
a	Na < Ne	nf=na
b	Na = Ne	nf=na=ne
c	Ne < Na < (1.2)(Ne)	nf=ne
d	Na > (1.2)(Ne)	nf=na

where

Na = the actual caseload,

Ne = the estimated caseload,

na = the sample size derived by substituting Na into the appropriate equation in figures 3-2, 3-3, or 3-4,

ne = the sample size based upon Ne, and

nf = the final minimum sample size.

3121.2 DESIRED SAMPLE SIZE. You may perform more reviews than allowed by the minimum sample size. You may want these, for example, to provide additional data for corrective action or to obtain greater reliability for subdivision analysis. The "desired sample size" refers to a sample size that is larger than the minimum sample size in order to provide additional reviews.

3121.3 TARGET SAMPLE SIZE. The minimum sample size, and the desired sample size, refer to the number of reviewable cases to be selected--that is, cases you are able to complete plus those you drop as Not Completed. (This is equivalent to the number of cases selected minus those you drop as Not Subject to Review.) The "target sample size" is the projected number of cases you must select in order to produce the number of reviewable cases specified by the minimum sample size (or the desired sample size).

To calculate the target sample size, multiply the minimum sample size (or desired sample size) by a factor to account for cases that are out of scope (i.e., will eventually be coded as Not Subject to Review). This overselection factor should not include reviews that will be coded as Not Completed since they are already contained in the minimum sample size (and the desired sample size).

Figure 3-6

EXAMPLE OF SAMPLE SIZE CALCULATIONS

A <u>Calculation of minimum sample size based on the estimated caseload</u>. (See 3121.1.)

Ne = 45,000 (estimated caseload subject to review)

For a non-reduced active sample, the minimum sample size is calculated as follows:

ne = 300 + [.042(45,000-10,000)]
ne = 1770 (estimated minimum sample size)

B <u>Calculation of the target sample size</u>. (See 3121.3.)

d = .04 (overselection factor)
ne(1+d) = 1770(1+.04) = 1841 (target sample size)

C <u>Calculation of the end-of-year minimum sample size</u>. (See 3121.1(B).)

Na = 55,000 (actual caseload subject to review)
Na > 1.2 * Ne because 55,000 > 1.2(45,000) = 54,000

na = 300 + [.042(55,000-10,000)]
na = 2190 (minimum sample size based on <u>actual</u> caseload)

nf = na = 2190 (final minimum sample size)

D <u>Calculation of the required sample size and the completion rate</u>. (See 3121.4.)

2200 cases are actually selected and 60 are found to be Not Subject to Review while 2110 were completed. (30 reviews were coded as Not Complete.)

The maximum of (2200-60) and 2190 is 2190 (required sample size).
2110/2190 = 96.35 % (completion rate)

E <u>Augmented sample size to provide for reserve-pool cases</u>. (See 3121.6.)

It is decided at the beginning of the period that a reserve pool of 10 percent will be selected.

1841 * .10 = 184 (additional cases planned for reserve pool)
1841 + 184 = 2025 (total sample to be drawn)

Part B of figure 3-6 presents an example of the calculation of the target sample size.

3121.4 REQUIRED SAMPLE SIZE. All selected cases that are reviewable must be completed. Consequently, we adjust your regressed error rate upward if your completion rate is less than 100 percent. Your completion rate is calculated by dividing the number of completed reviews by the "required sample size", which is the larger of (1) the number of cases selected that are reviewable or (2) the end-of-year minimum sample size, n_f[12].

Part D of figure 3-6 presents an example of the calculation of the required sample size and the corresponding completion rate.

3121.5 DEMONSTRATION-PROJECTS AND SSA CASES. You may want to select additional cases (in excess of the target sample size corresponding to your minimum sample size) to compensate for cases involved in demonstration projects or processed by the Social Security Administration. (The applicable demonstration projects are those we have authorized and have determined will significantly modify the rules for determining households' eligibility or allotment level.) The reason for considering the selection of additional cases is that demonstration and SSA cases are not used by FNS in the computation of error rates, but (if they are otherwise subject to review) are counted toward your required sample size in calculating your completion rate[13]. Consequently, an increase in your sample size may be desirable to produce more precise error-rate estimates.

3121.6 RESERVE-POOL SAMPLE. As a cushion against possible undersampling (see 3410), you may wish to select a reserve pool. This is a separate sample from which a supplemental sample can be drawn if undersampling occurs. The size of this pool is arbitrary but should be large enough to allow for any potential undersampling. The reserve pool sample may be selected separately but is usually drawn concurrently with the main sample. All cases in the same stratum must have an equal probability of being selected for the reserve pool. If only certain types of cases are candidates of inclusion in the reserve pool, then every case in a given stratum will not have an equal probability of being reviewed.

All reserve-pool cases in the same stratum must be drawn with the same probability of selection. This means that the size of each month's pool must be in the same proportions as the frame and the main sample; for example, if the caseload increases 10 percent from the previous month, then the size of the reserve pool for the second month must also be 10 percent larger. As explained in 3122, this is usually accomplished by selecting the reserve pool systematically using a constant interval.

12 7 CFR 275.11(d).

13 7 CFR 275.11(g).

If, during the year, you determine that a reserve pool is no longer necessary, you need not select a reserve-pool sample for subsequent months. (You must then also readjust your sampling intervals). This could occur if you determine that undersampling will not take place, or if you have implemented a correction for undersampling and do not anticipate the need for another one.

Part E of figure 3-6 presents an example of the augmentation of the target sample size to provide for reserve-pool cases.

3121.7 ALLOCATION OF STRATIFIED SAMPLES. If you are selecting a stratified sample, you must decide how you will allocate your sample among the strata. Your major consideration should be the expected precision of the stratified error-rate estimate. As explained in Exhibit E, the precision of any alternative design must be equal to or greater than the precision for a simple-random sample of size equal to your minimum sample size based on your estimated average-monthly caseload.

If you allocate your sample proportionally with respect to stratum caseloads, you produce a self-weighted sample. Then (1) weighting of State-calculated error rates (see Exhibit D) is not necessary, and (2) the equal-or-better-precision requirement is satisfied. For proportional allocation, calculate the sample size in each stratum by multiplying the total target sample size by the ratio of the stratum's frame size to the total frame size—that is,

$$n_i = n(N_i/N) \quad . \qquad (3\text{-}2)$$

Allocation that is not proportional necessitates the calculation of weighted error-rate estimates (see Exhibits D and H, attached). Moreover, a disproportionate allocation may need a larger total sample size to satisfy the equal-or-better-precision requirement (see Exhibit E). There are many textbooks to which you can refer that deal with the topic of sample allocation in stratified sampling (e.g., Cochran's Sampling Techniques and others listed in the Appendix.)

Once you have chosen the type of allocation and have determined the desired sample size in each stratum, you can calculate a target sample size for each stratum. As in unstratified sampling, the stratum-specific target sample size compensates for reviews that will be coded as Not Subject to Review. Also as in unstratified sampling, you can augment the target sample size in each stratum to provide reserve-pool cases by stratum.

3122 SAMPLING INTERVALS FOR SYSTEMATIC SAMPLING.

3122.1 GENERAL. After you have determined appropriate sample sizes, you must calculate sampling intervals for those samples to be selected systematically. In general, this is done by dividing the estimated caseload by the desired sample size (N/n). This is also applicable when computing the sampling intervals for individual strata in a stratified sample.

3122.2 WITHOUT RESERVE-POOL SUBSAMPLING. If you do not want a reserve pool, calculate the sampling interval (I) as follows[14]:

$$I = N_e' / n_t \ , \qquad (3\text{-}3)$$

where N_e' is your estimated total frame size, and n_t is your target sample size (see 3121.3). N_e' and n_t must both be twelve-month totals, or they must both be monthly averages. Figure 3-7 (part A) presents an example of the calculation of a sampling interval in which there is no reserve-pool subsampling.

3122.3 WITH RESERVE-POOL SUBSAMPLING. If you do want a reserve pool, equation (3-3) defines the effective interval, I_e. Use the following formula to calculate the actual interval, I_a, which you will use to select cases (reserve plus non-reserve) from the sampling frame.

$$I_a = N_e' / (n_t + n_p) \qquad (3\text{-}4)$$

where N_e' and n_t are defined above and n_p is the size of the reserve pool. N_e', n_t, and n_p must all be twelve-month totals, or they must all be monthly averages. After you have selected cases from the sampling frame with interval I_a, place into the reserve pool those cases obtained by subsampling the selected cases with the following subsampling interval:

$$I_p = I_e / (I_e - I_a) \ . \qquad (3\text{-}5)$$

Part B of figure 3-7 presents an example of sampling-interval calculations in which there is a reserve pool.

3122.4 ROUNDED VS. DECIMAL INTERVALS. In calculating sampling intervals, the quotients will obviously not always be integers. You can either round your results down to the nearest integer or retain one decimal place in the resulting figures depending upon the situation:

A You can round down (i.e., truncate) sampling intervals I, I_e, and I_a. This insures a sample size greater than or equal to the sample size that would result from using the exact interval. As an example, the calculations in Part A of Figure 3-7 truncate the resulting sampling interval.

B Retain one decimal place in the computation of I_p since much smaller sample sizes are involved. Part B of Figure 3-7 presents an example of a decimal sampling interval. 3123 and 3222 A,

---

14 This is equivalent to

$$I = N_e / n_r \ , \qquad (3\text{-}4')$$

where $N_e = N_e'(1-d)$ is the estimated subject-to-review caseload, n_r is the minimum sample size (see 3121.1) or desired sample size (see 3121.2), and d is the estimated out-of-scope rate. (N_e and n_r must both be twelve-month totals, or they must both be monthly averages.)

respectively, describe how to choose decimal random starts and how to select cases with decimal sampling intervals.

Figure 3-7

EXAMPLE OF SAMPLING-INTERVAL CALCULATIONS

(This example is a continuation of the calculations in figure 3-6.)

A Without reserve-pool subsampling. (See 3122.2)

$Ne = Ne' * (1-d)$
$(45{,}000 * 12) = Ne' * (1-.04)$
$Ne' = 540{,}000 / .96$
$Ne' = 562{,}500$ (estimated total frame size)

$I = Ne' / n_t$
$I = 562{,}500 / 1841$
$I = 305.54$, truncated to 305 (interval)

B With reserve-pool subsampling. (See 3122.3)

$I_e = I = 305$ (effective interval)

$I_a = Ne' / (n_t + n_p)$
$I_a = 562{,}500 / (1841 + 184)$
$I_a = 277.78$, truncated to 277 (actual interval)

$I_p = I_e / (I_e - I_a)$

$I_p = 305 / (305 - 277)$
$I_p = 10.9$ (interval for the reserve pool)

3123 RANDOM STARTS FOR SYSTEMATIC SAMPLING. Random starts for systematic sampling must be based on random numbers that are uniformly distributed over the range of the sampling interval. For example, if the sampling interval is 287, the numbers 1, 2, 3, etc., up through 287 must have equal probability of being selected as a random start. Similarly, if the sampling interval is 7.4, the numbers 0.1, 0.2, 0.3, etc., up through 7.4 must have equal probability of being selected as a random start.

The required number of random starts is determined by your method of sequentially numbering cases for systematic sampling. For example, if you assign "1" to the first case in the first month but assign 1+t to the first case in subsequent months, where t is the "tail count" of previous month's cases following the last sampled case, then you will need only one random start for the year. On the other hand if you assign "1" to the first case in each month, then you will need a random start for each month. The required number of random starts can also be effected if you change your sampling interval in the middle of the year. This may require you to either determine an additional random start or to replace some of your random starts (see 3410).

The numbers used to determine random starts must be either selected from a table of random digits or generated by a random-number algorithm executed in a computer or calculator. (Exhibit I, attached, contains a table of random digits plus describes some of the methods for generating random numbers in a computer program.) Alternative sources of random starts--such as throwing dice, flipping a coin, thinking of a number, drawing numbers from a hat, etc.--are not acceptable.

3130 SAMPLING PLANS.

3131 SAMPLING DOCUMENTATION. Your sampling plan describes and justifies your choice of sample design and sampling parameters[15]. In addition to the sampling plan, there are two other sources of information that describe your sampling procedures: supporting documents and the records of your sampling activities. Examples of supporting documents are (1) record layouts for computerized sampling frames and (2) flowcharts, source code, and operating instructions for associated computer programs. Examples of sampling-activity records are listings of your monthly samples and microfiche or hard-copy listings of your sampling frames. Acceptable documentation also includes the storage of the sampling frames on computer tapes written at the time of sample selection. Retain all supporting documents and sampling-activity records for at least three years[16]. These materials must, with sufficient notice, be available for our review (See 3420.). In addition, any sampling documentation associated with claims for or against the federal government must be retained until three years after the date of fiscal closure. This is the date the claims for or against the federal government have been liquidated [17].

3132 WRITING GUIDELINES. Include in your sampling plan a complete description of each of the items listed in 3133. Both technical and non-technical personnel will be reading your sampling plan. Consequently, provide illustrative examples of complicated procedures. Your sampling plan should stand alone as a single document and not require any supporting materials in order to be understood. We recommend you use the organizational format presented in 3133. You may, however, use a different format as long as all items are addressed and are easily identified.

3133 SAMPLING PLAN CONTENTS. In general, your sampling procedures and sampling parameters must conform to the principles of probability sampling and satisfy the demands of practicality. A non-integrated plan shall address all items listed in Exhibit J, attached. An

---

15 7 CFR 275.11(a) and 275.11(b)(4).

16 7 CFR 275.4.

17 7 CFR 272.1(f).

integrated plan shall also address all of these items and, in addition, include the following:

A A thorough description of the integration scheme, e.g., Modified Replacement B, explaining in detail how you will compile the sampling frames for each domain, what is the estimated number of sample cases that you will need in each subclass, and how you will replace cases;

B Actual or estimated caseloads, sample sizes, and actual out-of-scope rates for each stratum of the integration scheme; and

C An explanation of your correction for oversampling or undersampling in the event of an overage or shortfall in the caseload of a particular domain.

If you are changing your sample design from non-integrated to integrated, describe the impact of integration on the organization and management of the QC process--from the selection of the sample to the training of review staff to comparison of administrative costs with and without integrated sampling.

3134 SUBMISSION OF SAMPLING PLANS. Prior to the beginning of each year, send us either (1) a new plan, (2) a change to a previously-approved plan, or (3) a statement that a previously-approved plan will remain in use with no changes. If your plan is integrated, submit copies concurrently to the regional offices of those programs involved in the integration; e.g., Food Stamps, AFDC, and Medicaid. These submissions each year shall include the needed waiver statement if you have opted for a reduced active sample size (see 3121.1 and Exhibit G)[18].

If you are submitting either a minor change to a non-integrated plan or a statement that a sampling plan will remain in use without change, submit it at least 30 days prior to implementation. All other types of sampling-plan submissions--integrated plans, new plans, or major changes in a plan--are to be submitted at least 60 days prior to implementation[19]. Generally speaking, changes in sampling parameters (e.g., sample sizes, sampling intervals ,etc.) constitute minor changes, whereas changes in sampling procedures (e.g., moving to an integrated design or a change in frame construction) constitute major changes. For other types of changes, call us to determine if the change is major or minor. For minor changes, it is not necessary for you to send us your entire sampling plan--you may send us a description of only the minor changes. For major changes, however, send us your entire revised sampling plan.

18 7 CFR 275.11(a)(2).

19 7 CFR 272.2(e)(4).

3135 APPROVAL OF NON-INTEGRATED SAMPLING PLANS. After you have submitted your plan (in the manner described in 3134), we will evaluate it and either approve it or disapprove it. Sometimes it will lack sufficient information, however, for us to do either, in which case we will request you to provide additional information. Frequently, we will conditionally approve your plan contingent on your providing requested clarification or making certain modifications in the plan. We will inform you if your plan is approved (conditionally or unconditionally), disapproved, or if we need additional information. Do not implement your sampling plan until we have approved it.

3136 APPROVAL OF INTEGRATED SAMPLING PLANS. The approval of integrated plans is similar to that of non-integrated plans (see 3135), except that each respective regional office reviews your plan. A joint letter from the respective program's regional offices will be sent to you advising whether the integrated plan is approved, disapproved, requires additional information, or is approved for only one program. If a decision cannot be reached because additional information is needed or because of other factors, an interim reply will be sent advising you of the reason for the delay and when a final decision may be expected. The integrated design may be implemented only upon final approval of the respective regional offices.

3200 MONTHLY ACTIVITIES.

3210 FRAME CONSTRUCTION.

3211 GENERAL. FNS regulations address the appropriate composition of the Food Stamp sampling frames[20]. (See also 2100.) You must construct two Food Stamp sampling frames: one for the year's active cases, the other for its negative actions. These must be updated and samples selected from them each month. These may take the form of computer files, hard-copy listings, or batches of documents (e.g., application forms, household issuance records, termination notices, etc.). A sampling frame may be composed of one physical list, or a combination of many.

Though the sampling frames are updated every month, the data sources and processing procedures should remain the same from month to month. Describe these in your sampling plan (see 3130). Before using a data source or processing procedure for the first time, investigate the resulting sampling frame to insure that all appropriate Food Stamp households or actions have a known, nonzero, chance of selection. You can examine the computer files by (1) producing a hard-copy listing of the file (by "dumping" the file or by running the sample selection program against the file with an interval and random start both equal to one) and by (2) reading the computer program used to update the frame. All frames shall be retained, in their original order and composition, as directed in 3131 and 3420.

20 7 CFR 275.11(e),(f),and (g).

3212 CASES AND ACTIONS TO BE INCLUDED.

A Active Cases.

1 Inclusion Requirement. The sampling frame for active cases must contain all households that received Food Stamp benefits for any month during the year. The household must have been certified prior to the end of the month for which the associated benefits were issued.

2 Examples. The following are specific examples of households which should be included:

a Those certified but which did not receive benefits because of 100 percent recoupment;

b Those certified through expedited service;

c Those that received benefits via a manual issuance medium;

d Those that received the initial allotment of their certification period during the current sample month;

e Those which received a retroactive issuance, provided they were certified prior to the end of the month for which those benefits were issued. (These can be included in the sampling frames for either the month in which they were issued or the month for which they were issued. They are reviewed, however, with respect to the month for which they are issued. The State is not expected to include allotments for months which are more than one month removed from the month of issuance.)

f Households that were issued benefits under the rules of a demonstration project approved by FNS, or were certified through the Social Security Administration.

This is not meant to be an exhaustive list, and it assumes that the households satisfy the inclusion requirement.

3 Data Sources. You can construct Food Stamp sampling frames from a variety of data sources. These include:

a Participation - Those households which redeemed an ATP or were given food coupons directly. The rate of reviews listed in error is usually lower, but such lists are often not completely available until after the sample month;

b Issuance - Those households which were given an ATP or food coupons. Creation of this frame is not delayed as long as with participants, but a higher listed-in-error rate exists due to nonparticipation;

c Certification - Households which were approved for participation in the program (i.e., certified eligibles). The majority of this frame is available by the first day of the sample month, but many reviews will be dropped when the issuances are held and/or cancelled and when ATP's are not redeemed. Duplication usually is not a problem.

Select data sources for Food Stamp sampling frames on the basis of timeliness, completeness, and your administrative capabilities (e.g., consideration should be given to the 95-day time frame for completion and transmission of data). Any one source of data for constructing your sampling frames may not represent the entire target population. When this occurs, develop additional lists to supplement your main source of sampling units.

B <u>Negative Actions</u>.

1 <u>Inclusion Requirement</u>. The sampling frame for negative cases must contain all actions taken to deny an application or terminate a household prior to the end of its certification period.

The frame(s) may be constructed based on the action/decision date or on the effective date. (The review date of that action must be determined based upon Handbook 310.)

2 <u>Examples</u>. The following are specific examples of action which should be included:

a Actions that shorten the certification period;

b Households whose voluntary withdrawal request results in a termination;

c Each action taken against the household <u>that takes effect</u>, regardless of how many (i.e., multiple actions);

d Denial of a recertification;

e An action taken to deny or terminate a case due to noncooperation, death of all household members, or because the household moved from the State;

f Households denied or terminated under the rules of an FNS approved demonstration project, or whose application for Food Stamps was denied by the Social Security Administration; and g Negative actions on cases that continue to receive benefits pending a fair hearing.

3213 Out-of-Scope Cases and Actions. You should try to exclude from your sampling frame those cases that at the time of frame construction you know to be out of scope. (Those out-of-scope cases you fail to exclude, you code as Not Subject to Review if they are sampled.) You may exclude the following cases from your sampling frame if during frame construction you can determine if a case belongs to the specified category:

A Active Cases[21].

1 Households receiving benefits under a disaster certification authorized by FNS. This does not include households which are otherwise eligible and receiving benefits under regular certification standards.

2 A household in which all members have died or moved out of State before the review could be completed.

3 Households under investigation for an Intentional Program Violation (IPV) that (1) have already been referred to your fraud investigation unit and an investigation has been scheduled to begin within 5 months, (2) are currently under active investigation, or (3) have a pending administrative or judicial IPV hearing[22].

4 A household receiving only restored benefits for the sample month.

5 Those households which did not receive Food Stamp benefits. This includes those that were issued an ATP but did not redeem it[23]. Similarly, those instances where food coupons are returned as undeliverable by the Postal Service are also Not Subject to Review. (This does not apply to coupons returned by the client.)

---

21 7 CFR 275.11(f)(1). See Handbook 310 for a complete discussion of Not Subject to Review cases.

22 FNS Handbook 310, section 315.

23 An exception is that a household that had all or part of its allotment recouped to repay a prior Food Stamp overissuance is subject to review even if they did not cash an ATP.

B Negative Actions[24].

1 Those cases that are closed due to the normal expiration of the certification period.

2 A household that withdrew its application prior to agency's determination of eligibility.

3 A household in which all members have died or moved out of State subsequent to the selected action.

4 Actions associated with a transfer of category (e.g., PA to NA) or county responsibility.

5 An action to reduce benefits.

6 A household which was denied benefits through disaster certification proceedings.

C Suspensions. Actions to suspend a household are to be excluded from both the active and the negative frames.

3214 AVOIDING FRAME PROBLEMS. (See 2130 for sampling-concepts discussion.) You should make every effort to avoid sampling frame problems— including, but not limited to, those listed below:

A Periodic Order. If you sample systematically, your frame must not be periodic in order (i.e., having cyclical variation) in a manner such that the cycle is a multiple of the sampling interval. Such frames, however, are very rare.

B Duplicate Sampling Units. You must subject an active household to sampling and review only once for each month it receives benefits. (Because of retroactive benefits, an household may appear on a monthly frame more than once, but each appearance must represent a different month.) A particular household may have more than one negative action in a month, but each action must be included in the sampling frame only once.

In order to avoid the duplication of active households or negative actions, you should be aware of how such duplication can occur. The following are some examples which lead to multiple chances of selection:

1 All types of issuances are listed on the frame. For example, a given household may receive its initial allotment, then supplemental benefits, and then a replacement for the first issuance (which was lost). If you do not restrict the

24 7 CFR 275.11(f)(2). See Handbook 310 for a complete discussion of Not Subject to Review cases.

sampling frame to initial allotments (including retroactive issuances), duplication will exist.

2 An issuance may be temporarily held and then released later in the same month. If the frame is comprised of daily lists of issuances, such a situation could result in duplication.

3 In a state that selects an integrated sample, a household may be receiving only Food Stamp benefits at the beginning of the month but is then certified eligible for AFDC benefits in the middle of the same month. This household may be construed as both a "Food Stamp Only" and an integrated case for the same month, leading to possible duplication.

4 Some agencies may list a negative action once for each program (FS, AFDC, Medicaid, etc.) that is affected. Thus the action will have more than one chance of selection unless the frame is ordered in a way to allow the selector to combine sampling units.

5 A case should be listed only once for each negative action that is taken. If there is not enough information on the sample output to identify which action was chosen, each action still has only one chance to be selected, but may have more than one opportunity to be reviewed.

C Missing Sample Units. Not only must each household/action not have more than one chance to be selected and reviewed, but it should have exactly one chance. Each and every element in the universe must be represented in the frame. None should be omitted. This frequently requires frames which supplement the main source of sampling units.

Some common examples of types of cases which are overlooked are:

1 Households whose benefits are issued manually;

2 Cases which have their certification period shortened rather than terminated;

3 Retroactive issuances where the action to certify was made before the end of the benefit month; and

4 Decisions to deny or terminate that are not processed until the following month, or later in the same month.

D Clustered Sampling Units. If your sampling frame contains clustered sampling units, you must either (1) treat your sample as a cluster sample or (2) repeat the occurrence of the clustered sampling units in your sampling frame such that each occurrence corresponds to a unique population element. For example, assume a sampling frame consisted of sampling units defined to be participating households issued any benefits in the current month.

Thus, if a given household was issued benefits in the current month both for the current month and for a previous month, the household would appear only once in the sampling frame (because of the definition of the sampling unit). In some cases, however, the benefits for the previous month are subject to review (e.g., household is certified in the previous month but not issued the previous month's benefits until the current month), so that both allotments for such cases would be in the target population. If, however, the benefits only for the current month are reviewed, then reviewable previous-month benefits issued in the current month would never have a chance to be reviewed. To avoid this, you must provide for such cases to be represented twice in the sampling process, once for each month.

3215 CORRECTING FRAME PROBLEMS. It may be possible, under some circumstances, to mathematically correct for duplication without resampling. The possibility of bias from not representing some segment of the universe, however, can only be removed by the subsequent selection of a sample from that subpopulation (unless it can be shown that the missing segment has the same distribution of errors and allotments as the remainder of the population). Careful examination and frequent monitoring of the sampling frames is a necessity.

3220 SAMPLE SELECTION.

3221 GENERAL. Each month you must update your sampling frames for active and negative cases and select a random sample from each. In selecting each sample you must give all cases in the same stratum an equal chance of being selected[25]. Once you sample a household, you cannot replace it with a substitute[26]. For example, you cannot replace a household that is inconvenient geographically by another household, even if you substitute randomly. Such substitution can bias your sample. If you select a household for the active sample more than once in a year, review it for each monthly sample in which it appears. If you select a household for the negative sample more than once, review it for each separate denial or termination.

You must ensure the local office does not have prior knowledge of the cases scheduled for review. The reason is this may result in the local agency, intentionally or unintentionally, treating sampled cases in a special manner. This would cause your sample to be unrepresentative of the caseload and your estimated error rates to be biased. Consequently, you must take special precautions to ensure that sampled cases are not known to local offices prior to the case reviews.

---

25 The sample selection for cluster sampling (see Exhibit B) can be an exception to this statement.

26 The non-substitution rule applies to the final active or negative case samples--not to the intermediate overlap-domain samples in integrated sampling, for which the random replacement of cases between sampling frames is allowed.

3222 METHODS. There are only two different ways you can sample: you either sample individual cases (element sampling), or you sample groups of cases (cluster sampling). The remainder of this section describes the two most common methods of element sampling--systematic sampling and simple-random sampling--and describes one method of cluster sampling. Though there are only a few sample-selection methods, they can be incorporated into a number of different sample designs. For example, systematic sampling can be incorporated into stratified, unstratified, or integrated sample designs. The following sample-selection methods can be used to select not only your monthly samples but also other types of samples such as those for correcting for under- or oversampling (see 3410) or for replacing cases in integrated designs (see 2250).

A Systematic Sampling. (See 2230 for sampling-concepts discussion.) Systematic sampling requires that your sampling frame consists of one or more lists of cases and that it can be sequentially numbered. If you lack such lists or cannot easily sequence them, you should consider random digit sampling (see 3222 C) or cluster sampling (see Exhibit B).

Providing that your sampling frames are in suitable form, we generally recommend systematic sampling (either stratified or unstratified) for the selection of your monthly samples. We recommend systematic sampling because:

1 It is relatively easy to administer--especially for manual sampling;

2 It yields unbiased error-rate estimates, except in a few rare situations in which cases having similar probabilities of error appear in the sampling frame at equally spaced intervals; and

3 It yields a monthly sample size proportional to the monthly caseload.

Though we generally recommend systematic sampling, you may have certain conditions or constraints that dictate the use of alternative methods. Consequently, you may propose in your sampling plan alternative methods for sample selection.

The parameters for systematic sampling are the sampling interval and random start. 3122 and 3123 describe how to determine appropriate values for these parameters. 2230 describes systematic sample selection when the sampling interval is a whole number, such as 3 or 430.

When the sampling interval is a decimal number, such as 3.3 or 5.8, there are two sample selection methods to choose from. The easier method is the "accumulation method". This method generates two sequences of numbers. One is an unrounded sequence in which the first number is the random start and each successive number is obtained by adding the sampling interval to the immediately preceding number. The second sequence is a rounded sequence obtained by rounding the first sequence to the nearest whole number; it specifies the sequence numbers of the sampled cases. For example, assume the sampling interval is 4.6 and the random start is 1.7. Then the two sequences are as follows:

1.7	6.3	10.9	15.5	20.1	24.7	29.3	33.9
2	6	11	16	20	25	29	34

A second, and more involved, method is to use "replicated sampling". This uses k separate samples, each with a sampling interval given by

I_r = sampling interval for each sample replicate

$$= k * I \qquad (3\text{-}6)$$

where

I = the original sampling interval, and

k = the smallest integer such that multiplying k times I yields an integer.

For example, if I = 4.6, then k = 5, since five is the smallest integer that when multiplied by 4.6 gives an integer result (23). Thus I_r = 23. Each of the five samples requires a unique random start between 1 and 23. Assume these are obtained from a random number table as 5, 8, 11, 13, and 20. Then the first 15 sampled cases have the following sequence numbers:

5	8	11	13	20
28	31	34	36	43
51	54	57	59	66

B Simple-Random Sampling. (See 2220 for sample-concepts discussion.) The advantages and disadvantages of simple-random sampling depend on whether you select the sample manually or if a computer program selects it automatically.

1 Manual Selection. Except for very small samples, manual selection of a simple-random sample is more time consuming than manually selecting a systematic-random sample of the same size. The reason for this is the amount of time required to obtain each random number and to then check whether the corresponding case has already been sampled. Also, you must thoroughly document the source for all the random numbers so that we can reconstruct the sample in our annual reviews of your sampling procedures (see 3420). For these reasons, manual selection of a simple-random sample is allowed only for small samples such as reserve-pool sampling to correct for undersampling (see 3410 C) or replacement

operations in integrated sampling (see 2250). The advantage of simple-random sampling in these situations is that the desired number of cases to be sampled will be achieved exactly; whereas for systematic sampling the number of sampled cases may differ by one case from the desired sample size, depending on the random start and any rounding of the sampling interval.

2 Computer Selection. A computer program capable of selecting a simple-random sample must be able to generate random numbers. If you have access to a library function or already-written subroutine that does this (see Exhibit J), then the programming effort to automate simple-random sampling is about the same as that for systematic sampling[27].

There are two different approaches to automating the selection of a simple-random sample. The "fixed-sampling-fraction approach" requires a random-number generator that produces a real-value random number uniformly distributed over the interval between 0.0 and 1.0. Execute the random number generator for each case in the sampling frame. If the resulting random number is less than or equal to your sample's sampling fraction, then select the corresponding case to be in the sample; otherwise, the case is not included in the sample. In this approach the sample size is a random number with expected value equal to the product of the sampling fraction times the total number of cases in the sampling frames.

The "fixed-sample-size approach" has two sampling parameters: the number of cases in the sampling frame (denoted N) and the number of cases to be sampled (denoted n). Use this approach when both n and N are known—such as in corrections for under- or oversampling and in replacement operations for integrated sampling— and you want the sample to contain exactly n cases. Some subroutine libraries contain a subroutine that will take N and n as arguments and return n distinct integers uniformly distributed between 1 and N. If you lack such a subroutine, use the following algorithm[28] to sample exactly n cases:

Step 1: Set variable TN = n and variable TD = N. Set the "case-pointer" to point at the first case in the sampling frame.

---

27 State in your sampling plan the name and source of the computer algorithm you use to generate random numbers.

28 From Computer Simulation Techniques by Thomas Naylor, et. al., Wiley (1968), pp. 109-111.

Step 2: Calculate $T = TN / TD$. Generate a real-value random number that is uniformly distributed between 0.0 and 1.0. (Regenerate this number each time you execute Step 2.)

Step 3: If the random number from STEP 2 is less than or equal to T, then go to Step 4. Otherwise, do not include the case pointed to by the case-pointer in the sample and go to STEP 5.

Step 4: Include the case pointed to by the case-pointer in the sample and decrease the value of TN by 1.0.

Step 5: If $TN = 0.0$ or the case pointed to by the case-pointer is the last case in the sampling frame, then exit. Otherwise, reset the case-pointer to point at the next case in the sampling frame, decrease the value of TD by 1.0, and go back to STEP 2.

If you select simple-random samples for the monthly samples, you must maintain the same sampling fraction from month to month, unless you are stratifying by time. This is easily accomplished if you use the fixed-sampling-fraction approach. If, on the other hand, you use the fixed-sample-size approach, then you must calculate n by multiplying the constant sampling fraction by the monthly caseload. Using the same value of n from month to month is allowed only if successive monthly caseloads are exactly the same.

C Random Digit Sampling. Random digit sampling is a special case of cluster sampling: instead of sampling individual cases, you sample group-identification numbers, each of which is associated with a group of cases[29]. Use the following equation to determine the number of group-identification numbers that must be randomly selected each month:

$$m / M = n / (12 N) \qquad (3\text{-}7)$$

where

m = number of group-identification numbers to be sampled each month,

M = total number of possible group-identification numbers,

29 Use formula (B2), Exhibit B, to determine the precision of random digit sampling. If you use the variance formula for simple random sampling, you will in general underestimate the sampling variance because the variability in the size of the groups is a component of the variance.

n = desired number of cases to be reviewed during the year, and

N = average monthly caseload subject to review.

The same group number can be selected for more than one month, but the subsequent months selections must be completely random. To prevent prior knowledge, the selected group numbers cannot be non-randomly repeated from month to month.

A common application of random digit sampling uses a four-digit group-identification number and associates individual cases with groups using the last four digits of the head-of-household's social security number. (If the head of household does not have a social security number, you must randomly assign a four-digit number to the case[30].) Then $M = 10{,}000$ and

m = number of four-digit number to be randomly selected each month

$= 10{,}000\ n/\ (12\ N)$

$= 833\ n\ /\ N$ (3-8)

For example, if your annual number of reviews is 1200 and your average monthly caseload is 72,000, then each month you should randomly select (833) (1200 / 72,000) = 14 numbers.

For random digit sampling, we especially want to emphasize the following: (1) you must ensure that bias does not arise from local agencies' prior knowledge of the selected group-identification numbers, and (2) you cannot make substitutions for sampled cases. If all groups contain approximately the same number of cases[31], you do not have to weight your data to calculate error rates.

3230 MONTHLY REPORTING REQUIREMENTS.

3231 GENERAL. Each month report to us two types of information: the findings from your individual QC reviews and a summary of your sampling and review activities[32]. For the latter, complete Form FNS-248 ("Status of Sample Selection and Completion"), which we describe in detail in 3232. Send us the FNS-248 report so that we receive it no later than 105 days after the end of the sample month[33].

30 7 CFR 273.6 requires all individuals in a Food Stamp household to have a social security number.

31 A rule of thumb is that no group should differ in size from the average size of all the groups by more than 10 percent.

32 7 CFR 275.21.

33 7 CFR 275.21(c).

To report the individual findings for your active cases, use the Integrated Quality Control System software to transmit them to our National Computer Center at Kansas City. For negative cases, send us listings of your negative cases at least monthly indicating case dispositions and findings. Also, include lists of dropped (not completed and not subject to review) negative cases. All cases selected in a sample month shall be disposed of and the findings reported to us within 95 days from the end of the sample month.

For each case that remains pending 95 days after the end of the sample month, you must send us a report that includes an explanation of why the case has not been disposed of, documentation describing the progress of the review to date, and the date by which it will be completed. If we determine that this report does not justify the case's pending status, the case shall be considered overdue. Depending upon the number of overdue cases, we may find your QC system to be inefficient or ineffective and as a result suspend or disallow a portion of your Federal share of administrative funds[34].

3232 FORM FNS-248 SUMMARY REPORT. The Form FNS-248 you submit each month tracks the progress of your sample selections and case completions over the course of the year. Exhibit K, attached, contains the November 1989 revision of Form FNS-248 and the instructions for its completion.

3300 END-OF-YEAR ACTIVITIES.

3310 DETERMINATION OF THE ACTUAL OUT-OF-SCOPE RATE. At the beginning of the year, you use the projected out-of-scope rate, d, to calculate the target sample size (see 3121.3) and the predicted subject-to-review caseload (see 3110). At the end of the year, you use the actual out-of-scope rate to calculate the actual subject-to-review caseload (see 3320).

At the end of the year, calculate the actual out-of-scope rate as follows:

A Unstratified Samples.

1 Formula.

$$d = \text{out-of-scope rate}$$

$$= \frac{n(NSR) - n(deselect)}{n(select) - n(deselect)} \qquad (3\text{-}9)$$

where

34 7 CFR 275.21(b)(4).

n(NSR) = number of cases you disposed during the year as Not Subject to Review for any reason (including cases deselected in a correction for oversampling),

n(deselect) = number of cases you deselected during the year in any corrections for oversampling, and

n(select) = number of cases selected from the sampling frame, from a reserve pool, or in any supplemental sampling to correct for undersampling. Exclude cases you put into a reserve pool but did not withdraw from the pool to correct for undersampling. Include cases that you initially selected then later deselected to correct for oversampling.

2 Example. Prior to an adjustment for oversampling, you had selected 500 cases. Of these 500 cases, 50 were disposed as Not Subject to Review (NSR) because they were out of scope. The adjustment for oversampling deselected 20 cases from the 500 sampled cases. You changed the disposition codes for these 20 cases to NSR. (Unless the case was already NSR, and then no change was made.) At the end of the year, you had selected 1100 cases. (This includes the 20 deselected cases.) Of these 1100 cases, 90 had disposition codes of NSR. Thus,

$$d = (90 - 20) / (1100 - 20) = 0.065 .$$

B Proportionally Stratified Sample. Ignore strata and apply equation (3-9) to all of your data for the year.

C Disproportionately Stratified Samples. Use equation (3-9) to calculate the actual out-of-scope rate for each stratum.

3320 DETERMINATION OF THE ACTUAL SUBJECT-TO-REVIEW CASELOAD. At the end of the year, your actual subject-to-review caseload determines your minimum sample size if it is less than your projected caseload or more than 20 percent greater than the projected caseload (see 3121.1 B). Also, you will need actual subject-to-review caseloads if you use the direct approach (see 3110) to predicting caseloads for future years.

At the end of the year, use one of the following formulas to calculate the actual subject to review caseload:

A Unstratified Samples.

N = actual average-monthly caseload subject to review

$$= N'(1-d) \qquad (3\text{-}10)$$

where

N' = average-monthly frame size calculated from 12 months of exact counts of the number of cases in each month's sampling frame, and

d = the out-of-scope rate calculated at the end of the year (see 3310).

Substituting the formula for d in equation (3-9) into equation (3-10) gives the following alternative equation for calculating the actual subject-to-review caseload:

N = actual average-monthly caseload subject to review

$$= I\ [\ n(\text{select}) - n(\text{NSR})\ ]\ /\ m \qquad (3\text{-}10')$$

where

$$I = (m\ N')\ /\ [\ n(\text{select}) - n(\text{deselect})\ ]\ , \qquad (3\text{-}11)$$

(m N') = aggregation over m months of monthly frame sizes,

m = 12,

and n(select), n(NSR), and n(deselect) are defined in 3310. If you are sampling systematically and have not selected cases from a reserve pool, then in place of using equation (3-10) to calculate a value of I, you can use the following to obtain I:

1 If you are employing reserve-pool subsampling, I is the effective interval (see 3122.3).

2 If you have corrected for over- or undersampling by adjusting your sample (see 3410), I is the modified sampling interval in effect after the last sample adjustment.

3 If neither situation 1 nor 2 apply, then I is your sampling interval calculated at the beginning of the year (see 3122.2).

B Proportionally Stratified Samples. Ignore strata and apply equation (3-10) or (3-10') to all of your data for the year.

C Disproportionately Stratified Samples. Use equation (3-10) or (3-10') to calculate the actual caseload in each stratum. If you use equation (3-10'), m is the number of months the stratum is in effect. For example, if stratum 1 is October through February and stratum 2 is March through September, then m = 5 for stratum 1 and m = 7 for stratum 2. On the other hand, if stratum 1 is public assistance (PA) cases for the entire year and stratum 2 is non-PA cases for the entire year, then m = 12 for both strata.

Except in the case of stratification by time (see 3414), you do not have to calculate a weighted or combined average-monthly caseload. For the case of stratification by time, calculate a weighted average-monthly caseload by weighting together the stratum-level average-monthly caseloads with weights of m/12, where m is the number of months the stratum is in effect.

3330 COMPLETION OF FORM FNS-247. Form FNS-247 summarizes the data obtained from your active and negative samples over the course of the year. Exhibit L, attached, contains the February 1990 revision of Form FNS-247. (Prior to its revision, this form contained four parts numbered 247.1 through 247.4. Part 247.1 we revised and renamed as the (revised) Form FNS-247, which no longer includes parts 247.2, 247.3, and 247.4. Though you can obtain automated reports for Parts 247.2, 247.3, and 247.4 from the IQCS software, we no longer require you to send them to us. At the end of the year, complete Part Form FNS-247 according to the instructions in Exhibit L.

We must receive your completed Form FNS-247 no later than 105 days after the end of the fiscal year—that is, by January 13[35].

In completing Form FNS-247, account for all cases selected in your active and negative samples. Include cases that are completed, not completed, and not subject to review. In addition, break out the annual totals for those cases involved in approved demonstration projects or processed by the Social Security Administration [36]. (There is no column on Form FNS-247 in which to place these additional totals. Consequently, you will have to place that information at the bottom of the Form FNS-247. Annotate the additional totals with a footnote.)

If you have a disproportionately stratified sample, enter stratum data in the additional columns provided on the Form FNS-247. A separate Form FNS-247 will have to be prepared if your sample contains more than four strata. Enter the stratum identifier and stratum interval in the column headings provided. In addition, calculate an entire-sample entry for lines 2 and 4 (end of year minimum sample sizes) and a weighted entry for line 12 (the negative-case error rate). Annotate each Form FNS-247 as to what information it is summarizing.

If your sample is disproportionately stratified, record the data for individual strata in separate columns. Do this for all lines containing four columns. Enter stratum identifiers in the column

---

35 7 CFR 275.21(d) and (e). Prior to the February 17, 1984 regulatory change, the Form FNS-247 consisted of four separate parts: Parts 247.1, 247.2, 247.3, and 247.4. Part 247.1 was renamed FNS-247. Parts 247.2, 247.3, and 247.4 were made optional by this rulemaking.

36 Item 27 of the Integrated Review Schedule contains the case-classification code, which codes cases processed by the Social Security Administration and cases involved in approved demonstration projects as 2 and 3, respectively. See Section II of the Integrated Manual for AFDC, Adult, Food Stamp and Medicaid Quality Control Reviews.

headings. If the number of strata exceeds four, use additional forms. For lines containing only one column, enter the requested entire-sample value.

3340 END-OF-YEAR RECONCILIATION. At the end of the year, we will send you data listings for cases you have transmitted to our National Computer Center in Kansas City (NCC-KC). These listings will show for completed cases those data entries important for error-rate estimation. Though what you transmit during the year is your official data, it is important to reconcile any differences between the provided listings and your own records. For example, your records may show more cases than do the listings of the data at NCC-KC; this may be because there are some cases you have not yet transmitted to NCC-KC.

Once a case has been subjected to federal subsampling the protected fields cannot be changed. There are some fields, however, that can and must be corrected if they are incorrect because of their impact on error-rate estimation. Depending on your sample design, the correctness of some or all of the following fields may be crucial for a valid estimation of the error-rate for your State:

A Sample Month and Year (field 3 of the Integrated Review Schedule),

B Stratum Code (field 4), and

C Case classification (field 27).

Part of the reconciliation process is to ensure that all of your transmitted cases contain the appropriate coding on the fields mentioned above. Another part of the process is to identify cases that (as a result of post-transmission data changes) have erroneous findings, such as a correct case with an error amount.

3400 OTHER ACTIVITIES PERFORMED DURING THE YEAR.

3410 SAMPLING CORRECTIONS. In preparing your sampling plan, you must predict various sampling parameters for the coming year. These include caseload sizes, out-of-scope rates, and overlap proportions. Such predictions permit you to plan for an expected number of case reviews. Later during the year, however, you may find the resulting number of reviews to be very different from what you had planned.

If this occurs, consider making a sampling correction. A "correction for undersampling" increases your sampling rate[37]. You should consider making this type of correction if you suspect without it your

37 See 2210 for the definition of "sampling rate". For systematic sampling the sampling rate is the reciprocal of the sampling interval.

annual number of reviews will be less than your end-of-year minimum sample size (see 3121.1). If such a shortfall does occur, your completion rate (defined in 3121.4) will be less than 100 percent, and we will consequently adjust your regressed error rate upward[38]. Depending upon the magnitude of the shortfall, we may find your QC system to be inefficient or ineffective and consequently suspend or disallow a portion of your Federal share of administrative funds[39].

A "correction for oversampling" decreases your sampling rate. You may want to perform a correction for oversampling if you find yourself in the following situation:

A You suspect without a change in sampling rate your annual number of reviews will be considerably greater than the larger of (1) your end-of-year minimum sample size or (2) your desired annual number of reviews, and

B A portion of the projected excess number of reviews will occur in the months that follow.

In deciding whether to make a sampling correction, you compare your annual number of reviews with your minimum sample size. At the end of the year, you, of course, know these values. At mid-year, however, you must project them. 3411 describes the associated calculations.

There are two different methods for making sampling corrections: sample adjustment and stratification-by-time. A sample adjustment changes the sampling rate in all 12 months of the year. Stratification-by-time changes the sampling rate only in those months following your decision to make a sampling correction. 3412 and 3413 describe sample adjustment; 3414 describes stratification-by-time. For both methods, 3415 and 3416 describe associated reporting and data management requirements.

Because a sampling correction changes one or more sampling parameters, it changes your sampling plan. Hence, we must approve your proposed sampling correction prior to its implementation. For a sample adjustment, call us first for approval. If we approve it, then send us a description of your new sampling parameters and your adjustment of samples in preceding months. Stratification-by-time, on the other hand, requires demonstration of equal-or-better precision. Consequently, if you are stratifying by time, send us a written proposal 30 or more days prior to implementation.

3411 DETERMINING THE NEED FOR A SAMPLING CORRECTION. The key factor in determining whether you should make a sampling correction is the magnitude and direction of your error in predicting the average-monthly

38 7 CFR 275.23(e)(6)(iii).

39 7 CFR 275.11(c)(2), requirement to correct excessive undersampling; 7 CFR 276.4, suspension/disallowance of administrative funds.

caseload subject to review. Consequently, you must first calculate an updated estimate of the average-monthly caseload and then determine the effects of any caseload change on the expected and minimum sample sizes.

3411.1 UPDATING ESTIMATED AVERAGE-MONTHLY CASELOAD. Each month update your estimate for the year of the average monthly caseload subject to review. Use the following formula:

N_u = updated estimate for the entire year of average monthly caseload subject to review

$$= [(m)(N1) + (12-m)(N2)] / 12, \qquad (3\text{-}12)$$

where

m = number of preceding months,

N1 = average-monthly subject-to-review caseload calculated by applying equation (3-10) or (3-10') (see 3320) to the preceding m months, and

N2 = projected average monthly caseload subject to review, averaged over remaining 12-m months.

If you believe the m preceding months are representative of the entire year, use instead the following formula:

$$N_u = N1. \qquad (3\text{-}13)$$

3411.2 DETERMINING THE EFFECTS OF CASELOAD CHANGES.

A Unstratified Systematic Samples: Use of "Quick Test". Except for a few rare situations (described below), you can for systematic samples quickly determine if you shouldn't make a sampling correction. This "quick test" is based on the relationship between the beginning-of-the-year estimate of average monthly caseload, N_e, and the updated estimate, N_u. Perform the quick test as follows:

1 Determine the applicable row of the table in figure 3-1 (see 3114) based on (1) the type of case--active or negative; (2) your decision about reducing the active sample size; and (3) the interval containing both N_e and N_u. (If N_e and N_u are in different intervals, then the quick test cannot be used, and you should make use of the work sheet described in 3411.2 B.)

2 Determine the applicable column of the table based on the relationship between N_e and N_u. For example, if N_u is larger than N_e, use the column labeled "over-prediction".

3 A positive number at the intersection of the selected row and column indicates a SURPLUS. That is, the number of reviews

is expected to exceed the minimum sample size. Consequently, you should not correct for undersampling, and you may want to consider a correction for oversampling.

4 A negative number at the intersection of the selected row and column indicates a DEFICIT. That is, the number of reviews is expected to be less than the minimum sample size. Thus, you should correct for undersampling. (NOTE: This result assumes that the sampling interval is exactly equal to the beginning-of-the-year estimate of the average monthly caseload divided by the initial minimum sample size per month. If the sampling interval is smaller than this because of (1) rounding down to the nearest integer or (2) your adding extra cases to the denominator of the sampling-interval formula, then a correction for undersampling may not be needed even though the quick test indicates a DEFICIT.)

You can estimate the magnitude of the SURPLUS or DEFICIT by multiplying the applicable entry of table 3-1 by the number of percentage points of caseload under- or over-prediction--that is, by

$$\text{absolute percentage error} = |100(N_u - N_e)/N_e|^* . \qquad (3\text{-}14)$$

Depending on the magnitude of the SURPLUS or DEFICIT and how many more months remain to be sampled, make a decision on whether to correct for under- or oversampling. It may be preferable to postpone making a correction until later in the year when a trend may be more definite.

The quick test will give incorrect results if (1) your sample is disproportionately stratified or not systematic, (2) your sampling interval is based on a desired sample size that is larger than the minimum sample size, or (3) you fail to calculate the interval in the manner described in 3122.

B General Procedure: Use of Work Sheet.

The work sheet in Exhibit M, attached, presents a general procedure to determine the effects of a caseload change. This procedure can be used for any type of sample.

3412 SAMPLE ADJUSTMENT OF UNSTRATIFIED SAMPLES.

A General. A sample adjustment changes the sampling rate in all 12 months of the year. To correct for undersampling, it increases the sampling rate; whereas to correct for oversampling, it decreases the sampling rate. If sampling is systematic, a sample adjustment corrects for undersampling by decreasing the sampling interval and corrects for oversampling by increasing the sampling interval.

---

* $|x|$ is the symbol for the absolute value of x. The absolute value of a number is always positive. For example, $|-3| = +3$.

When you adjust a sample during the year, you (1) change the sampling rate for following months and (2) modify the samples from previous months. The type of modification of the samples from previous months depends on whether you are correcting for under- or oversampling. When correcting for undersampling, you modify the previous-months samples by selecting additional cases from these months. When correcting for oversampling, you deselect some of the previous months' cases. For a sample adjustment at the end of the year, there are no following months, but you do modify the sample for all previous 12 months.

No matter when you do a sample adjustment--during the year or at the end of the year--you must redetermine your monthly sampling parameter(s). Part B explains how to do this. You then modify your samples for the previous months. Part C(1) explains how to do this when correcting for undersampling; Part C(2) for oversampling.

B Determine New Sampling Parameter(s). You must redetermine the monthly sampling parameter(s), even for an end-of-year sample adjustment. Do this the same way as you did at the beginning of the year except for the following changes:

1 Use updated values (see 3411) for both (1) minimum sample size and (2) average-monthly caseload subject to review, instead of the predicted values in the sampling plan; and

2 If appropriate, use updated estimates of sampling-frame characteristics, such as proportions of overlap or listed-in-error cases.

3 If you are sampling systematically and are adjusting the sample during the year, you must reselect some or all of random starts for subsequent months in the following situations:

a If adjusting for oversampling, reselect all random starts for following months. The reason is that your revised sampling interval is larger than the previous interval, and all previously selected random starts were uniformly distributed only up to the previous sampling interval.

b If adjusting for undersampling, you need reselect only those following-months' random starts that are larger than the revised sampling interval.

c If in the absence of a sample adjustment you would have needed no following-months' random starts (because of retaining prior-months' "tail counts"), then you will need to select only one random start (uniformly distributed over the revised sampling interval) to determine the first sampled case following the sample adjustment.

C <u>Modify Samples From Previous Months</u>.

1 <u>Correcting Previous Months for Undersampling</u>.

a <u>Sampling Frames for Supplemental Sampling</u>. To correct the previous months for undersampling, obtain additional cases by selecting a supplemental sample for the previous months. You can use one of two methods: replicated sampling or subsampling from a reserve pool. In <u>replicated</u> sampling you perform all sampling at the time of sample adjustment by selecting the sample from one of the following sampling frames:

i All cases previously exposed to sampling (called the "all-previous-cases sampling frame"), or

ii All cases previously exposed to sampling that were not selected (called the "non-selection sampling frame").

Alternatively, you can <u>subsample</u> from a reserve pool to select the supplemental sample. This consists of two sampling phases. You perform the first phase every month by selecting "reserve cases", which are cases you do not review the first time you sample them. Instead, you set them aside for use in a second phase of sampling. This occurs at the time of sample adjustment--when you obtain the supplemental sample by subsampling the reserve pool consisting of all reserve cases from previous months. You then review for the supplemental sample only the reserve cases you have subsampled.

If you are sampling systematically, 3122.3 explains how to determine the sampling intervals for putting cases <u>into</u> a reserve pool. Either systematic or simple-random sampling can be used to remove cases <u>from</u> a reserve pool. (See 3222A and 3222B.)

Generally, replicated sampling is easier than reserve-pool subsampling if (1) you use a computer for sample selection and (2) you retain in computer-readable form all the sampling frames for previous months. Otherwise, you may find reserve-pool subsampling to be the easier method.

Describe in your sampling plan your frame for the supplemental sample. (In reserve-pool subsampling, the reserve pool is the supplemental sampling frame.) The selection of the supplemental sample must be random. Also, every case in a given stratum of your supplemental sampling frame must have equal probability of being selected.

b Supplementing a Systematic Sample. If you are (1) selecting the additional cases by replicated sampling and (2) using systematic sampling in the original and supplemental samples, calculate the interval for the supplemental sample from one of the following formulas:

i For sampling from the all-previous-cases sampling frame,

I(supp) = sampling interval for supplemental sample

$$= (I)(I')/(I-I') \tag{3-15}$$

ii For sampling from the non-selection sampling frame,

$$I(supp) = (I-1)(I)/(I-I') \tag{3-16}$$

where

I = previous-months sampling interval, and

I' = new, following-months sampling interval.

Generally, the sampling interval for the supplemental sample will be large. Consequently, it is not necessary to use a decimal interval for selecting the supplemental sample.

c General Procedure. When you are not able to use the above sampling-interval formulas, use the following procedure for determining the target sample size for the supplemental sample:

Step 1: Determine the subsequent-months sampling parameters(s), even if this is an end-of-year sample adjustment. (See Part B.)

Step 2: Use results from Step 1 to compute the subsequent-months sampling rate.

Step 3: Multiply result from Step 2 by the number of cases in the all-previous-cases sampling frame to estimate number of cases in the previous months that would have been selected had the new sampling parameter(s) been in effect.

Step 4: If you made a sample adjustment correcting for oversampling in the previous months, subtract the number of deselected cases from the total number of cases selected for review in the preceding months. Otherwise, the result of this step is the total number of cases selected in the preceding months.

Step 5: Subtract the result from Step 4 from the result from Step 3 to obtain target sample size for the supplemental sample.

The above procedure gives the target sample size of the supplemental sample regardless of the supplemental sampling frame and the type of sampling.

d Cases Selected Twice. There are two situations in which it is possible for the supplemental sample to select a case that has been previously selected. One of these situations is when you select the supplemental sample from the all-previous-cases sampling frame. Then it is possible to select the same case in both the original sample and the supplemental sample. Except in States with very small caseloads, this is extremely rare. If it does occur, however, keep the case in the supplemental sample with the same QC findings as in the original sample but change its review number for the supplemental sample. The reason for this is that retaining the case in the supplemental sample is statistically valid, but for purposes of data management every sampled case (in the same six-month period) must have a unique review number[40].

The other situation for possible double selection occurs when a correction for undersampling follows a correction for oversampling (see Part C.2). A case can be selected in the original sample, deselected in the oversampling correction, and then reselected in the undersampling correction. This is also extremely rare. If this should occur, however, keep the case in the supplemental sample and maintain its deselection from the original sample. (It is not necessary to change the review number to identify the case in the supplemental sample.)

2 Correcting Preceding Months For Oversampling. If you want to correct for oversampling, you can either (1) adjust your sample, which involves deselecting some previously selected cases, or (2) stratify your sample by time (see 3414). If you have completed the reviews for a large portion of your preceding-months' cases, we generally recommend stratification by time instead of sample adjustment because

40 The Integrated Quality Control Software only requires unique review numbers within each of the two 6-month periods: October through March and April through September.

the latter discards information for already-completed reviews. In integrated sample designs, however, if stratification by time has undesirable effects on AFDC and Medicaid QC activities, then sample adjustment would be the preferred method for correcting for oversampling.

If you are selecting your monthly samples systematically, use the following formula to calculate the deselection interval:

$$I(\text{deselect}) = I / (I - I') \qquad (3\text{-}17)$$

where

I = the previous-months sampling interval, and

I' = the new, following-months sampling interval.

You use the deselection interval to systematically sample from the set of all previously selected cases. (This includes completed, not completed, and not subject to review cases.) The cases chosen in this deselection sample you code as Not Subject to Review.

If you are not selecting the monthly samples systematically, you must calculate the number of cases to be deselected. You can use the general procedure described in the preceding section for determining the target sample size of the supplemental sample for an undersampling correction. This procedure will yield a negative number, which is the number of cases you should deselect using either systematic sampling or simple random sampling without replacement.

3413 ADJUSTING STRATIFIED SAMPLES. A sample adjustment of a stratified sample retains the same number of strata, but changes the stratum sampling rates. This is different from a stratification-by-time correction (see 3414), which increases the number of strata. The procedure for adjusting a stratified sample depends on a number of different factors. These include (1) the comparison of sampling rates across strata before adjustment (all-equal or not-all-equal), (2) comparison of stratum sampling rates after adjustment, and (3) the relationship between pre- and post-adjustment sampling rates. This latter factor arises from the following equation:

$$k_h = f'_h / f_h, \qquad (3\text{-}18)$$

where

f_h = the sampling rate for stratum h before sample adjustment, and

f'_h = the sampling rate for stratum h after sample adjustment.

If the k_h's are the same across all strata, the sample adjustment is one of "constant scaling"; otherwise, it is said to be of "variable scaling". The table in figure 3-8 assigns a situation number to each possible combination of the pertinent factors.

Figure 3-8

SITUATION NUMBERS FOR ADJUSTMENT OF STRATIFIED SAMPLES

sampling rates before adjustment	type of scaling	sampling rates after adjustment all-equal	not-all-equal
All-equal	Constant	Situation 1	Impossible
All-equal	Variable	Impossible	Situation 2
Not-all-equal	Constant	Impossible	Situation 3
Not-all-equal	Variable	Situation 2 (special case)	Situation 2

Here's how to handle each of these situations:

A Situation 1. Ignore strata. Adjust the sample as if it were unstratified. (In an undersampling correction, allocate the number of supplemental cases across strata proportional to each stratum's total frame size for prior months. In an oversampling correction, allocate the number of deselected cases across strata proportional to each stratum's sample size for prior months.)

B Situation 2. The sample adjustments for Situation 1 or Situation 3 possesses a possible drawback in integrated samples; that is, the constant scaling of sampling rates may adversely effect AFDC or Medicaid QC activities. Variable scaling may avoid this problem. To perform a Situation 2 adjustment, you must first reallocate the Food Stamp sample based on the updated minimum sample size (see 3411). This reallocation must satisfy the equal-or-better precision requirement (see Exhibit E) and the restraints that caused you to reject constant scaling. Then calculate following-month sampling parameters for each stratum and use these results to specify supplemental samples to be drawn from the previous months in each stratum in a correction for undersampling or to specify deselection in each stratum in a correction for oversampling. Because of the complexity of a Situation 2 adjustment, we don't recommend it unless both constant scaling and stratification-by-time (see 3414) are unacceptable.

C Situation 3. The subsequent-month sampling rates are given by the following equation:

$$f'_h = k\, f_h \,, \tag{3-19}$$

where

$$k = N_e\, n_u \,/\, (N_u\, n_e) \tag{3-20}$$

and f_h and f'_h are defined above. Thus, if you are correcting for undersamping k is greater than one, and if for oversampling, less than one. If you are sampling systematically, then

I_h' = new, following-months sampling interval for stratum h

$$= I_h/k \,, \tag{3-21}$$

where

I_h = previous-months sampling interval for stratum h.

Use these results in equations (3-15), (3-16), or (3-17) to specify supplemental samples to be selected in each stratum when correcting for undersampling or to specify deselection in each stratum when correcting for oversampling.

3414 STRATIFICATION BY TIME. The second way of correcting for under- or oversampling is to stratify by time. This designates the cases in months following the correction to be in a different stratum from the cases in previous months. For example, if you make a sampling correction in March, then October through February are one stratum, and March through September, a second stratum. If you make a second sampling correction in June, then October through February are one stratum; March through May, a second stratum; and June through September, a third stratum.

The purpose of stratification by time is to allow you to use different sampling rates in different strata. Therefore, in the stratum that follows the sample correction, increase the sampling rate to correct for undersampling and decrease it to correct for oversampling. If you are sampling systematically, this corresponds to decreasing the sampling interval to correct for undersampling and increasing the sampling interval to correct for oversampling.

The resulting stratified sample design must satisfy the equal-or-better-precision requirement described in Exhibit E. As is discussed in E340, if the true error rates are the same in different time strata, then using stratification by time produces a loss of precision. Thus, in order to maintain precision, your minimum sample size increases if you stratify by time. This increase is fairly small, however, except when the stratum allocations are extremely disproportionate. In the case of correcting for oversampling, the increase in the minimum sample size resulting from stratification by time is usually less than the number of cases that would have been deselected if you had adjusted the previous-months' sample instead of stratifying by time.

If you are sampling systematically, the following formula yields the appropriate sampling interval for the stratum containing the months subsequent to the sampling correction:

$$I_p = \text{sampling interval for the post-correction stratum}$$

$$= (12-m)\ N2\ /\ n_p \qquad (3\text{-}22)$$

where

n_p = sample size for the post-correction stratum that is needed to satisfy the equal-or-better-precision requirement (See E340.)

and m and N2 are already defined in 3411.1.

3415 REPORTING REQUIREMENTS. The following sampling corrections entail additional end-of-year reporting activities (See 3330.):

A Supplemental Sampling From a Reserve Pool. If you correct for undersampling by randomly selecting cases from a reserve pool, then at the end of the year you must tell us how many cases were put into the reserve pool during the year and how many cases were selected from the reserve pool. This reporting requirement can be satisfied by your preparing a separate FNS-247 report for the reserve pool cases. If cases were put into a reserve pool but none were selected for a sampling correction, then reporting of information about the reserve cases is not required.

B Stratification by Time. Complete the FNS-247 report in accordance with the directions for a stratified sample design.

3416 DATA MANAGEMENT REQUIREMENTS. The following sampling corrections have associated data management requirements:

A Sample Adjustment To Correct for Oversampling. Code deselected cases as Not Subject to Review. (For cases you have already transmitted as Completed or Not Completed, provide us a list of those you have deselected, and we will change their dispositions.)

B Stratification by Time. Use separate stratum codes to distinguish those months prior to the sampling correction from those months following the sampling correction.

3420 PREPARATION AND PARTICIPATION IN FNS REVIEWS OF SAMPLING PROCEDURES. Annually, we will review your procedures for sampling, reporting and data management[41]. We will determine if they satisfy your sampling plan, FNS policy memoranda, regulations CFR 275.11 through 275.13, and standard statistical practices. The review will consist of discussions

41 7 CFR 275.3(c)(1)(iv).

between you and the FNS reviewer about these procedures and his/her examination of related documents.

The major portion of the review concerns your sampling of active and negative cases. We will examine your sampling procedures for the current year; that is, for the fiscal year containing the date of the review. Also, if we believe it necessary, we will review your sampling procedures for the preceding year.

A Preparation. In preparation for the review, you should be prepared to:

1 Discuss sampling procedures in general;

2 Reconstruct sampling procedures using the original sampling frames or using copies of the original sampling frames (such as copies stored on microfiche or magnetic tape);

3 Examine computer programs/flow charts which construct the sampling frame(s) and select the sample;

4 Review the methods used to estimate caseloads;

5 Outline procedures and controls used to obtain, edit, tabulate and report Quality Control data; and

6 Assemble any other material or data requested by the Regional Office.

B Assembly of Materials. Before the review, assemble the materials needed to document the above activities. These materials include:[42]

1 Original sampling frames for active and negative cases,

2 Active and negative sample lists,

3 Computer programs constructing the sampling frame(s) and selecting the sample,

4 Caseload estimation work sheets, and

5 Tracking logs.

C Review Agenda. Necessary staff should also be available for discussion of the sampling activities and documentation. A typical review agenda would incorporate the above points, optional brief entrance and exit conferences, and discussion of specific sampling problems.

---

42 7 CFR 275.11(a)(4) and 275.4(a), (c).

GLOSSARY

"Accuracy" means an opposite-sense description of bias. If an estimate's bias is large, the estimate is said to have low accuracy. If, on the other hand, the estimate's bias is small, the estimate is said to have high accuracy. The term "validity" is synonymous with "accuracy".

"Active case". See section 130 of Handbook 310.

"Actual out-of-scope rate" means the end-of-year ratio of the number of out-of-scope cases to the number of reviewable cases. It is calculated from equation (3-9), and for disproportionately stratified samples a separate calculation is performed for each stratum.

"Actual reviewable caseload" means the end-of-year estimate for the average-monthly caseload subject to review. It is calculated from the equations in 3320.

"Actual sampling interval". One way to select reserve-pool cases is to use two-phase systematic sampling: you first select a systematic sample and then systematically subsample it. Cases selected in both the (first-phase) sample and the (second-phase) subsample are reserve-pool cases; whereas, cases selected in the sample but not in the subsample are cases to be reviewed. The first-phase sampling interval is called the "actual sampling interval". It is calculated from equation (3-4).

"Actual subject-to-review caseload". See "Actual reviewable caseload".

"Allocation" means in stratified samples the mathematical breakdown of the overall sample size into the individual stratum sample sizes. See "Proportional allocation" and "Disproportionate allocation".

"Beginning-of-the-year minimum sample size" means for (unstratified) systematic or simple-random samples the minimum sample size based on the beginning-of-year estimate of the reviewable caseload and the minimum-sample-size-schedules in figures 3-2, 3-3, and 3-4. For an alternative sample design, the beginning-of-the-year minimum sample size is that sample size for the alternative sample design that yields the same predicted precision as the beginning-of-the-year sample size for a simple-random sample.

"Better precision" means smaller sampling variance.

"Bias" means the amount of the difference between the expected value of an estimate and the true value of the population parameter being estimated.

"Caseload weight" means in stratified samples the ratio of the annual reviewable caseload in an individual stratum to the annual reviewable caseload for the target population. If your sample is disproportionately stratified, you use caseload weights in estimating case-error rates and in calculating combined-ratio estimates of dollar-error rates.

"Cluster sampling" means the sampling units are clusters of several (or many) elements.

"Clustered sampling unit" means that a sampling unit contains more than one population element.

"Combined-ratio estimate" means a ratio estimate for a stratified sample in which both numerator and denominator are weighted sample means.

"Completion rate" means the ratio of the number of completed reviews to the required sample size.

"Confidence interval". See D500.

"Control data" means the information contained in a sampling frame for overlapping populations that indicates for each sampling unit the population(s) to which the sampling unit belongs.

"Cross-classification" means the determination of a sampled-unit's domain membership after sample selection.

"Desired sample size" means a sample size that is larger than the minimum sample size in order provide additional reviews.

"Direct approach" means a method for predicting the reviewable caseload that makes use of a historical data series of reviewable caseloads.

"Disjoint sampling frames" means, when sampling overlapping populations, two or more sampling frames containing no common sampling elements.

"Disproportionate allocation" means in stratified sampling the absence of all the stratum sampling fractions being equal.

"Disproportionate stratification". See "Disproportionate allocation".

"Dollar weight" means in stratified samples the benefit dollars issued to subject-to-review households in a given stratum divided by the benefit dollars issued to all subject-to-review households in the target population. If your sample is disproportionately stratified, you use dollar weights in calculating separate-ratio estimates of dollar-error rates.

"Domain" means a subset of the target population.

"Dropped case" means a case that you have disposed as either Not Subject to Review or Not Completed.

"Effective sampling interval". One way to select reserve-pool cases is to use two-phase systematic sampling: you first select a systematic sample and then systematically subsample it. Cases selected in both the (first-phase) sample and the (second-phase) subsample are reserve-pool cases; whereas, cases selected in the sample but not in the subsample are cases to be reviewed. The effective sampling interval is the interval you would use to obtain from a single-phase systematic sample the same number of cases to be reviewed and no reserve-pool cases. The effective sampling interval is calculated from equation (3-3).

"Elements" means the elementary units for which you seek information.

"Element sampling" means that each sampling unit contains no more than one element.

"End-of-year minimum sample size" means for an (unstratified) systematic or simple-random sample the minimum sample size based on (1) your actual reviewable caseload, (2) your estimated reviewable caseload, and (3) the formulas in figure 3-5. For alternative sample designs, see E330.

"Epsem" means "equal-probability selection method".

"Equal-probability selection method" means an element-sampling scheme in which every sampling unit has equal probability of being sampled.

"Estimate" means a specific value that results from a computation involving sample data.

"Estimated reviewable caseload" means your beginning-of-the-year prediction of the average-monthly caseload subject to review.

"Estimation error" means the difference between an estimate and the population parameter being estimated.

"Estimator" means the computational procedure used for obtaining an estimate.

"Expected value" means the average value of the estimates resulting from all possible samples that can be selected in accordance with your sampling plan.

"Final minimum sample size". See "End-of-year minimum sample size".

"Fixed-sample-size approach" means in simple-random sampling that the defining sample parameters are the size of the population and the exact size of the sample to be selected.

"Fixed-sampling-fraction approach" means in simple-random sampling that the defining sample parameters are the size of the population and the sampling fraction.

"Frame". See "Sampling frame".

"Integrated sample design" means a sample design in which some (or all) of the sampled units receive combined data collection for more than one population.

"Interval". See "Sampling interval". Also, see "Actual sampling interval" and "Effective sampling interval".

"Minimum sample size". See "Beginning-of-the-year minimum sample size" and "End-of-the-year minimum sample size". Also, see "Desired sample size", "Required sample size", and "Target sample size".

"Negative action". See "Negative case".

"Negative case". See section 130 of Handbook 310.

"Negative error rate" means the ratio of the number of invalid negative actions to the number of negative actions.

"Out-of-scope" means a description given to a sampling unit that is not a member of the target population but is, nevertheless, contained in the associated sampling frame. An out-of-scope case is one you dispose as Not Subject to Review for reasons other than its deselection to correct for oversampling.

"Overlap domain" means a set of elements which is common to two or more target populations.

"Overlapping sampling frames" means, when sampling overlapping populations, two or more sampling frames that contain some common sampling units.

"Overpayment error rate" means the proportion of all Food Stamp benefits issued in a fiscal year that are either (1) issued to households that fail to meet the program eligibility requirements or (2) overissued to eligible households.

"Parameter". See "Population parameter" and "Sampling parameter".

"Payment error rate" means the sum of the point estimates of the estimated over- and underpayment error rates.

"Population elements". See "elements".

"Population parameter" means a quantity calculated from data for all the elements in the target population. Examples of population parameters are population means and population totals.

"Precision" means an opposite-sense description of sampling variance. If an estimate's variance is large, the estimate is said to have low precision. If on the other hand, the estimate's variance is small, the estimate is said to have high precision.

"Proportional allocation" means in stratified sampling that all of the stratum sampling rates are equal.

"Proportional stratification". See "Proportional allocation".

"Random-digit sampling" means a special case of cluster sampling in which, instead of sampling individual cases, you sample group-identification numbers, each of which is associated with a group of cases.

"Random start" means in systematic sampling a random number that is uniformly distributed over the range of the sampling interval. If the sampling interval is an integer, the random start is an integer. If the random start is a decimal value, such as 3.4, then the random start is uniformly distributed over the values 0.1, 0.2, etc., up to the value of the sampling interval.

"Reliability". See "Precision".

"Replicated sampling" means the selection of two or more samples from the same target population. Some applications of replicated sampling are (1) as an alternative method for selecting a systematic sample using a decimal sampling interval and (2) in selecting a supplemental sample in order to correct for undersampling.

"Required sample size" means the larger of (1) the number of cases selected that are reviewable or (2) the end-of-year minimum sample size.

"Reserve pool" means a separate sample of Food Stamp cases from which a supplemental sample can be selected.

"Sample design" means a description of your sampling procedure(s) and the associated populations. A sample design need not specify the values for the sampling parameters but may instead describe how these values are to be determined.

"Sampled unit" means a sampling unit that has been selected to be in a specific sample.

"Sample-fraction approach" means a method for predicting the subject-to-review caseload for an overlap domain in which you (1) choose one of the associated populations and predict its caseload subject to review, (2) estimate (from sample data) the proportion of cases in the chosen population that are also in the overlap domain, and (3) multiply the predicted subject-to-review caseload for the chosen population by the estimated overlap proportion to produce a prediction of the subject-to-review caseload for the overlap domain.

"Sampling" is the selection of a subset of units to represent a larger set of units.

"Sampling error" is the component of estimation error that occurs when you collect data only for a sample and not the entire population.

"Sampling fraction" means the ratio of sample size to population size.

"Sampling frame" means the physical representation of the target population. A sampling frame is usually a list of sampling units. If there is no such list, a sampling frame is some equivalent procedure for identifying sampling units.

"Sampling-frame approach" means a method for predicting the average-monthly reviewable caseload in which you first make two predictions: a prediction of N_e', the estimated average monthly frame size; and a prediction of d, the out-of-scope rate. The product $N_e'(1.0 - d)$ then estimates the average-monthly caseload subject to review.

"Sampling-frame problems" means situations in which there is a lack of one-to-one correspondence between sampling units in the sampling frame and elements in the target population.

"Sampling interval" means in systematic sampling the reciprocal of the sampling rate.

"Sampling parameter" means a numerical quantity that characterizes a sampling procedure. The size of a sample is an example of a sampling parameter.

"Sampling plan" means the description of your sample design along with the specification of your sampling parameters.

"Sampling procedure" means a sequence of operations that produces a sample.

"Sampling rate". See "Sampling fraction".

"Sampling unit" means a distinct and identifiable unit of the target population containing an integer (possibly zero) number of elements. The collection of all sampling units corresponding to a target population must cover the whole of the population and they must not overlap, in the sense that every element in the population belongs to one and only one sampling unit.

"Sampling variance" means the average over all the samples that can be selected in accordance with your sampling plan of the squared difference between the sample estimate and the expected value of the estimate.

"Separate ratio estimate" means a ratio estimate for a stratified sample that is a weighted sum of the individual stratum ratio estimates.

"Simple-random sampling" means a method of sampling from a sampling frame of N elements in which any of the possible subsets of n distinct sampling units chosen from the sampling frame is equally likely to be a sample of size n.

"Standard error" means the square root of the sampling variance.

"Strata" means the disjoint and exhaustive sets of sampling units resulting from the stratification of a sampling frame.

"Stratification" means the partitioning of a sampling frame into disjoint and exhaustive sets of sampling units.

"Stratification by time" means, the creation of strata as a result of not using the same sampling rate for every sample selection.

"Stratified sampling" means a method of sampling from a stratified sampling frame in which a separate, independent sample of a specified size is selected from each stratum.

"Stratum" is the singular of "strata".

"Subclass" means a set of sampled units all of which as a result of cross-classification are assigned to the same domain.

"Subsampling" means the selection of a sample from a sample.

"Systematic sampling" means a method of sampling in which (1) the sampling frame consists of a sequenced list and (2) for a given random start the selected sampling units are those whose sequence numbers equal (to within rounding) either the random start or the random start plus an integer multiple of the sampling interval.

"Target population" means the set of all elements of interest.

"Target sample size" means the projected number of cases that you must select in order to produce the number of reviewable cases specified by your minimum sample size (or the desired sample size).

"Totally integrated frame" means, when sampling overlapping populations, a single sampling frame that contains all of the sampling units for all the target populations.

"Underpayment error rate" means the ratio of the value of benefits underissued to participating households to the total value of the benefits issued to participating households.

"Universe". See "Target population".

"Validity". See "Accuracy".

"Variance". See "Sampling variance". For the definition of a "case variance", see the definition of "variance" in section 130 of Handbook 310.

CLUSTER SAMPLING

B100 GENERAL. There are a few alternative sampling designs which you can opt for as a result of 7 CFR 275.11(a)(1). Cluster sampling is one of these and we discuss it in this exhibit because of our experience with it. Some initial considerations and an example of single-stage cluster sampling are provided to assist you in deciding whether a cluster sampling design is appropriate and feasible.

Cluster sampling is described as a division of the population into groups or clusters of elements with a subsequent selection of a random sample of these clusters. Consequently, the sampling unit is a cluster and not a population element. Although the population is divided into groups for both cluster and stratified designs, the latter draws a sample of elements from every group, whereas the former obtains elements only from a selected sample of groups.

Cluster sampling has advantages and disadvantages to your operations. The geographical proximity of Food Stamp households reduces reviewer travel time and related expenses. However, a potential loss of precision exists and, therefore, more total reviews may be required than would a simple random or systematic sample to maintain equivalent precision. Also, the upfront planning and the effort to maintain cluster sampling may use resources extensively.

For FNS to consider your cluster sampling design, you must describe the associated administrative logistics and the resources for performing the same. FNS will disapprove the cluster sampling plan if: 1) in our view, you lack sufficient resources to perform the associated administrative logistics, 2) the plan is not technically accurate, 3) your cluster sampling procedures cannot be verified by us, 4) your cluster sampling procedures are incompatible with the Federal regression methodology, or 5) the cluster sampling design fails the equal-or-better-precision requirement. If the cluster sampling design is integrated as well, then early consultation with the respective Regional Offices (Food Stamp Program, Aid to Families with Dependent Children (AFDC), Medicaid) is advisable.

B200 DEFINING CLUSTERS. Clusters may take any of a number of forms, e.g., project areas, counties or certification offices. The decision to use a particular type of cluster will depend on several factors:

A The cluster must be well defined. Every household subject to quality control review must belong to one and only one cluster.

B The number of households in each cluster must be known or a reasonable estimate must be available.

C Clusters should be relatively small. As a general rule, it is often best to construct a large number of relatively small, heterogeneous clusters.

D Clusters should be chosen to minimize the increase in sampling error caused by clustering. (For example, if you have large clusters, split them into smaller ones or subsample within the clusters.)

It is not necessary that clusters be defined identically everywhere, provided that each household belongs to one and only one cluster. In urban areas, clusters may consist of city blocks or groups of blocks. In rural areas, they may be defined by natural boundaries or counties. It is not necessary that every cluster be of equal size. In general, natural clusters will vary in size. Large variations will increase sampling variability, but some techniques are available to control variation in cluster size or the effects of this variation. For example, the more heterogeneous the clusters, the smaller the increase in sampling variability as a result of variations in cluster size.

B300 <u>SINGLE STAGE CLUSTER DESIGN.</u> This is a discussion of a single stage cluster design with equal probabilities of cluster selection. More sophisticated cluster designs and procedures for determining optimum cluster sizes exist and are described in the sampling textbooks listed in the Appendix.

Once appropriate clusters have been specified, a simple random sample of clusters is selected from a frame listing every cluster in the population. Estimates of the case and allotment error rates can be obtained from:

$$p = \sum x_i / \sum m_i \tag{B1}$$

with estimated variance:

$$VAR(p) = (1/n\bar{M}^2)\ (\sum (x_i - pm_i)^2)/(n-1) \tag{B2}$$

where: x_i = number of sample cases (dollars) in error in cluster i

m_i = number of sample cases (dollars) in cluster i

n = number of clusters selected in the sample

$\bar{M}$ = M/N = the average cluster size of the population

M = number of cases (dollars) in the population

N = number of clusters in the population

If information is unavailable to compute $\bar{M}$, it can be estimated by:

$$m = \sum m_i/n \tag{B3}$$

These formulas are interchangeable for case and allotment error rates. To compute case error rates, substitute the number of cases in error in each cluster, the total number of sample cases in each cluster, and the total number of households in the population for x_i, m_i, and M respectively. To compute allotment error rates, substitute the amount of the allotments issued in error within each cluster, the total value of the allotments issued within each cluster, and the total value of the allotments issued to the population for x_i, m_i and M respectively.

B400 EXAMPLE. In the following example we compare a cluster sample to the corresponding simple-random sample with respect to the variance of estimated case error rate.

Cluster	Number of Households (m_i)	Number of Households in Error (x_i)
1	40	12
2	60	10
3	20	3
4	25	10
5	30	5
6	30	7
7	35	2
8	25	5
9	40	6
10	15	9
11	10	1
12	30	6
13	25	12
14	50	5
15	45	7
16	15	2
17	30	6
18	25	7
19	25	5
20	20	10
21	30	10
22	40	7
23	35	8
24	15	5
25	40	10

$\sum m_i = 755$

$\sum x_i = 170$

$\sum x_i m_i = 5420$

$\sum m_i^2 = 26175$

$\sum x_i^2 = 1380$

The best estimate of the population proportion of cases in error is:

$$p = \sum x_i / \sum m_i = 170/755 = 0.225$$

To estimate the variance of p, first compute

$$\begin{aligned}\sum (x_i - pm_i)^2 &= \sum x_i^2 - 2p \sum x_i m_i + p^2 \sum m_i^2 \\ &= 1380 - 2(0.225)(5420) + (0.225)^2(26175) \\ &= 266.109\end{aligned}$$

Estimate $\bar{M}$ by $\bar{m}$, where

$$\bar{m} = \sum m_i/n = 755/25 = 30.2$$

Then,

$$\begin{aligned}VAR(p) &= (1/n\bar{M}^2)\ (\sum (x_i - pm_i)^2/(n-1) \\ &= 266.109/25(30.2)^2(25-1) \\ &= 0.000486\end{aligned}$$

If these data had been obtained from a simple random sample, the estimated variance would have been:

$$VAR(p) = pq/n = \frac{(0.225)(0.775)}{755} = .0002309$$

The ratio of $\dfrac{\text{VAR(p)cluster}}{\text{VAR(p)simple random}} = 0.000486/0.000231 = 2.1$

shows that the clustering scheme more than doubles the sampling error in this example. This loss of precision must be compensated by increasing the overall sample size.

INTEGRATED SAMPLE DESIGNS

C100 GENERAL. This Exhibit contains brief summaries of the sample designs in the Integrated Quality Control System (IQCS) sampling handbook published in December 1979. Each summary includes:

A A list of sampling parameters, and

B A summary table, which displays for each subclass the expected number of sample cases and the number and types of reviews to be performed.

Some of the summaries also contain an example. When an example is present, it is based on the corresponding example in the IQCS sampling handbook. You should read the corresponding material in the IQCS sampling handbook in order to completely understand the example.

C200 INFORMATION CONTAINED IN THE SUMMARY TABLES. The rows of the summary table list the five possible non-overlapping domains for single- or joint-program participation in Food Stamps (FS), Aid for Dependent Children (AFDC), or Medicaid. The columns of the table represent sampling frames or strata within a sampling frame. Above each column is a description of the frame or strata.

Each non-blank table cell displays two lines of information. The first line contains a label for the subclass and specifies the expected subclass size. The latter is usually a single symbol (same as in the IQCS sampling manual[1]) but may be a formula. In either case, the expected subclass size is placed at the end of the first line of information and is enclosed in parentheses.

The purpose of the subclass label is to provide a quick and easy way to refer to a set of cases. The label for the subclass assumes various forms depending on the sampling procedure that produces the subclass. If a subclass is all the cases selected from one stratum or one sampling frame, the label is a single number; e.g. "1", "2", etc. If the subclass results from cross-classification, the label is of the form "Ip"—e.g. "1a", "2c", etc.—indicating the association of cases in sample "I" with domain "p". If a subclass results from subsampling, the label is of the form "I.J"—e.g. "1.1", "1.2", etc.—indicating subsample "J" selected from larger sample "I".

The second line of information specifies the number and types of reviews to be performed. This information is enclosed in square brackets. If all cases in the subclass are to receive the same type of reviews, this is indicated as follows:

1 In the IQCS sampling handbook, all multiply-subscripted symbols comply with a common notational convention. The initial subscript indicates the sampling frame from which cases are selected. Non-initial subscripts indicate other programs for which the households are subject to review.

[FAM]	All cases reviewed for FS, AFDC, and Medicaid
[AM]	All cases reviewed for AFDC and Medicaid
[FM]	All cases to be reviewed for FS and Medicaid
[F]	All cases reviewed only for FS
[M]	All cases reviewed only for Medicaid
[Ø]	All cases not reviewed

On the other hand, if the subclass is to be randomly subdivided, with one set of cases receiving one type of review, and the remaining set of cases to receive another type of review, this is indicated in the following way:

[(n)A, (m)B]	n cases reviewed for program(s) A, and remaining cases reviewed for program(s) B

Some examples of this are the following:

[(n) FAM, (m) AM]	review n cases for FS, AFDC, and Medicaid, and review m cases for AFDC and Medicaid
[(n) F, (m) Ø]	review n cases for FS, and do not review m cases

C300 SUMMARIES. The following sections summarize the sample designs in the IQCS sampling handbook:

- C310 Replacement Method A--not utilizing the FS/Medicaid overlap
- C320 Replacement Method A--utilizing the FS/Medicaid overlap
- C330 Modified Replacement Method A
- C340 Replacement Method B
- C350 Modified Replacement Method B
- C360 Social Security Number Method
- C370 Sequential Random Sample Method
- C380 Disproportionate Stratified Method A
- C390 Modified Disproportionate Stratified Method A

When an example is included, the summary is presented on two pages facing each other. When there is no example, the summary is presented on a single page.

C310 REPLACEMENT METHOD A (NOT UTILIZING FS/MEDICAID OVERLAP).

C311 SAMPLING PARAMETERS.

n_a, n_f, n_m: AFDC, FS, and MAO (non-AFDC) sample sizes.
n_{af}, n_{ao} from cross-classification of n_a.
N_f = estimated average monthly caseload of subject-to-review FS cases.
N_{fa} = estimated average monthly caseload of subject-to-review FS cases that are receiving AFDC.
$n'_{af} = (N_{fa}/N_f)\ n_f$.
n_{fo}, see table note.

C312 SUMMARY TABLE[2].

Domain	AFDC cases	AFDC cases rcvng FS	FS cases not rcvng AFDC	Medicaid cases not rcvng AFDC
FS/AFDC/Medicaid	1a (n_{af}) [FAM]	2*		
AFDC/Medicaid	1b (n_{ao}) [AM]			
FS/Medicaid			3 (n_{fo}) [F]	4 (n_m) [M]
FS only			3, with above	
Medicaid only				4, with above

*If $n_{af} < n'_{af}$, then $n_{fo} = n_f - n'_{af}$ and Sample 2 consists of $n'_{af} - n_{af}$ cases, which receive only FS reviews; otherwise, $n_{fo} = n_f - n_{af}$ and Sample 2 is omitted.

2 The summary table depicts the description of Replacement Method A on pages 12-15 of IQCS sampling Handbook, which does not utilize the FS/Medicaid overlap. The example of replacement Method A on pages 15-16 of the handbook does, however, utilize the FS/Medicaid overlap. This modification is not reflected in the summary table.

C313 EXAMPLE. (Based on example on pages 15 and 16 of IQCS sampling handbook but does not utilize the FS/Medicaid overlap.)

$n_a = 602$ $n_f = 112$ $m = 291$

$n_{af} = 254$
$n_{ao} = 602 - 254 = 348$

$n'_{af} = (N_{fa}/N_f)n_f = 619 > n_{af}$
$n_{fo} = 1120-619 = 501$

Domain	AFDC cases	AFDC cases rcvng FS	FS cases not rcvng AFDC	Medicaid cases not rcvng AFDC
FS/AFDC/Medicaid	254 [FAM]	619-254 = 365 [F]		
AFDC/Medicaid	348 [AM]			
FS/Medicaid			* 501 [F]	*
FS only			*	291 [M]
Medicaid only				*

Total reviews = 1759

C320 REPLACEMENT METHOD A (UTILIZING FS/MEDICAID OVERLAP)[3].

C321 SAMPLING PARAMETERS.

n_a, n_f, n_m: AFDC, FS, and MAO (non-AFDC) sample sizes.
n_{af}, n_{ao} from cross-classification of n_a.
N_f = estimated average monthly caseload of subject-to-review FS cases.
N_{fa} = estimated average monthly caseload of subject-to-review FS cases that are receiving AFDC.
$n'_{af} = (N_{fa}/N_f)\ n_f$.
n_{fo}, see table note.
n_{fm} from cross-classification of n_{fo}.
n_{mf}, n_{mo} from cross-classification of n_m.

C322 SUMMARY TABLE.

Domain	AFDC cases	AFDC cases rcvng FS	FS cases not rcvng AFCC	Medicaid cases not rcvng AFDC
FS/AFDC/Medicaid	1a (n_{af}) [FAM]	2*		
AFDC/Medicaid	1b (n_{ao}) [AM]			
FS/Medicaid			3a (n_{fm}) [(n_{mf}) FM, ($n_{fm} - n'_{fmo}$) F]	4a (n_{mf}) [Ø]
FS only			3b ($n_{fo} - n_{fo}$) [F]	
Medicaid only				4b (n_{mo}) [M]

*If $n_{af} < n'_{af}$, then $n_{fo} = n_f - n'_{af}$ and Sample 2 consists of $n'_{af} - n_{af}$ cases which receive only FS reviews: otherwise, $n_{fo} = n_f - n_{af}$ and Sample 2 is omitted.

3 The example of replacement Method A on pages 15-16 of the IQCS Sampling Handbook utilizes the FS/Medicaid overlap. This modification of Replacement Method A, which can be used only if $n_{fm} \geq n_{mf}$, is reflected in the summary table in C322.

C323 EXAMPLE. (See pages 15 and 16 of IQCS sampling handbook.)

$n_a = 602$ $n_f = 1120$ $n_m = 291$

$n_{af} = 254$

$n_{ao} = 602 - 254 = 348$

$n'_{af} = (N_{fa}/N_f)n_f = 619$

$n_{fo} = 1120-619 = 501$

$n_{fm} = 40$

$n_{mo} = (.87)(291) = 253$

$n^m{}_f = 291 - 253 = 38 < n_{fm}$

Domain	AFDC cases	AFDC cases rcvng FS	FS cases not rcvng AFDC	Medicaid cases not rcvng AFDC
FS/AFDC/Medicaid	254 [FAM]	619-254 = 365 [F]		
AFDC/Medicaid	348 [AM]			
FS/Medicaid			~~40~~ [38 FM, 2F]	~~38~~ [Ø]
FS only			461 [F]	
Medicaid only				253 [M]

Total reviews = 1721

C330 MODIFIED REPLACEMENT METHOD A.

C331 SAMPLING PARAMETERS.

n_a, n_f, n_m: AFDC, FS, and MAO (non-AFDC) sample sizes.
n_{afm}, n_{am} from cross-classification of n_a.
n_{fma} from cross-classification of n_f.
n_{mf} from cross-classification of n_m.
n''_{fma}, n''_{fm}, n''_{fo}, see table note (*).
n''_{mf}, n_{mo} see table note (§).

C332 SUMMARY TABLE.

Domain	AFDC cases	FS cases not rcvng AFDC	Medicaid cases not rcvng AFDC
FS/AFDC/Medicaid	1a (n_{afm}) [FAM]	2a (n''_{fma})* †	
AFDC/Medicaid	1b (n_{am}) [AM]		
FS/Medicaid		2b (n''_{fm})* [FM]	3a (n''_{mf})§ ¶
FS only		2c (n''_{fo})* [F]	
Medicaid only			3b (n''_{mo})§ [M]

*If $n_{afm} < n_{fma}$, then n''_{fma}, n''_{fm}, and n''_{fo} are from cross-classification of n_f. Otherwise, from cross-classification of n'_f, which results from adjusting n_f based on a precision calculation.

†If $n_{afm} < n_{fma}$, then [(n_{afm}) Ø, ($n_{fma} - n_{afm}$) FM].
If $n_{afm} \geq n_{fma}$, then [Ø].

§If $n''_{fm} < n_{mf}$, then n''_{mf} and n''_{mo} are from cross-classification of n_m. Otherwise, from cross-classification of n'_m, which results from inflating n_m based on a precision calculation.

¶If $n''_{fm} < n_{mf}$, then [(n''_{fm}) Ø, ($n_{mf} - n''_{fm}$) M].
If $n''_{fm} > n_{mf}$, then [Ø].

C333 <u>EXAMPLE</u>. (See pages 22 through 24 of IQCS sampling handbook.)

$n_a = 840$ $n_f = 1200$ $n_m = 175$

$n_{afm} = 588$

$n_{am} = 252$

$n_{fma} = 560 < n_{afm}$

$n_{mf} = 53$

$n''_{fma} = n'_{fma} = 530$ $n''_{fm} = n'_{fm} = 37$ $n''_{fo} = n'_{fo} = 578$[4]

$n''_{mf} = n_{mf} = 50$ $n''_{mo} = n_{mo} = 125$

Domain	AFDC cases	FS cases not rcvng AFDC	Medicaid cases not rcvng AFDC
FS/AFDC/Medicaid	588 [FAM]	~~530~~ [Ø]	
AFDC/Medicaid	252 [AM]		
FS/Medicaid		37[4] [FM]	~~50~~ [~~37~~ Ø, 13 M][4]
FS only		578[4] [F]	
Medicaid only			125 [M]
Total reviews = 1593			

4 These results differ slightly from those on page 24 of the IQCS handbook. The reason for this is that the IQCS handbook fails to retain sufficient number of decimal places in intermediate calculations. Moreover, because of the assumption of equal population variances implicit in formulas on pages 18 and 19 of the IQCS sampling handbook, it can be shown mathematically that $n_f \leq n_{afm} + n'_{fm} + n'_{fo}$. The results in the IQCS sampling handbook do not satisfy this condition whereas the above results do.

C340 REPLACEMENT METHOD B[5].

C341 SAMPLING PARAMETERS.

n_a, n_f, n_m: AFDC, FS, and MAO (non-AFDC) sample sizes .
n_{af}, n_{ao} from cross-classification of n_a.
N_f = estimated average monthly caseload of subject-to-review FS cases.
N_{fa} = estimated average monthly caseload of subject-to-review FS cases that are receiving AFDC.
$n_{fa} = (N_{fa}/N_f)\ n_f$.
$n_{fo} = n_f - n_{fa}$.

C342 SUMMARY TABLE.

Domain	AFDC cases	FS cases not rcvng AFDC	Medicaid cases not rcvng AFDC
FS/AFDC/Medicaid	1a (n_{af}) [(n_{fa}) FAM, ($n_{af}-n_{fa}$)AM]		
AFDC/Medicaid	1b (n_{ao}) [AM]		
FS/Medicaid		2 (n_{fo}) [F]	3 (n_m) [M]
FS only		2, with above	
Medicare only			3, with above

5 Replacement Method B can be used only if $n_{af} \geq n_{fa}$. Also see footnote 2, above.

C343 EXAMPLE. (See pages 15 and 16 of IQCS sampling handbook.)

$n_a = 602$ $n_f = 1120$ $n_m = 291$

$n_{af} = 400$

$n_{ao} = 602 - 400 = 202$

$n_{fa} = (N_{fa}/N_f)n_f = 300 < n_{af}$

$n_{fo} = 1120 - 300 = 820$

Domain	AFDC cases	FS cases not rcvng AFDC	Medicaid cases not rcvng AFDC
FS/AFDC/Medicaid	~~400~~ [300 FAM, 100 AM]		
AFDC/Medicaid	202 [AM]		
FS/Medicaid		*	*
		820 [F]	291 [M]
FS only		*	
Medicare only			*

Total reviews = 1713

C350 MODIFIED REPLACEMENT METHOD B.

C351 SAMPLING PARAMETERS.

n_a, n_f, n_m: AFDC, FS, and MAO (non-AFDC) sample sizes.
n_{afm}, n_{am} from cross-classification of n_a.
n_{fma}, n_{fm}, n_{fo} from cross-classification of n_f.
n_{mf}, n_{mo} from cross-classification of n_m.

C352 Summary Table.

Domain	AFDC cases	FS cases not rcvng AFDC	Medicaid cases not rcvng AFDC
FS/AFDC/Medicaid	1a (n_{afm}) *	2a (n_{fma}) *	
AFDC/Medicaid	1b (N_{am}) [AM]		
FS/Medicaid		2b (n_{fm}) †	3a (n_{mf}) †
FS only		2c (n_{fo}) [F]	
Medicaid only			3b (n_{mo}) [M]

*The number and type of reviews performed in subclasses 1a and 2a are determined by the parameters n_{afm} and n_{fma}:

	Subclass 1a	Subclass 2a
$n_{afm}<n_{fma}$	[FAM]	[(n_{afm})Ø, ($n_{fma}-n_{afm}$)FM]
$n_{afm}>n_{fma}$	[(n_{fma}) FMA, ($n_{afm}-n_{fma}$)Ø]	[Ø]

†The number and type of reviews performed in subclasses 2b and 3a are determined by the parameters n_{fm} and n_{mf}:

	Subclass 2b	Subclass 3a
$n_{fm}\le n_{mf}$	[F]	[(n_{fm})Ø, ($n_{mf}-n_{fm}$)M]
$n_{fm}>n_{mf}$	[(n_{fm})FM, ($n_{fm}-n_{mf}$)F]	[Ø]

C353 EXAMPLE. (See page 27 of IQCS sampling handbook.)

$n_a = 840$ $n_f = 120$ $n_m = 175$

$n_{afm} = 588$

$n_{am} = 252$

$n''_{fma} = n_{fma} = 560 < n_{afm}$

$n''_{fm} = n_{fm} = 40$

$n''_{fo} = n_{fo} = 600$

$n''_{mf} = n_{mf} = 50 > n_{fm}$

$n''_{mo} = n_{mo} = 125$

Domain	AFDC cases	FS cases not rcvng AFDC	Medicaid cases not rcvng AFDC
FS/AFDC/Medicaid	~~588~~ [560 FAM, 28 AM][6]	~~560~~[6] [Ø]	
AFDC/Medicaid	252 [AM]		
FS/Medicaid		40 [FM]	~~50~~ [40 Ø, 10 M]
FS only		600 [F]	
Medicaid only			125 [M]
Total reviews = 1615			

6 These results differ from those on page 27 of the IQCS sampling manual, which use the reduced Food Stamp sample (Sample 2) from the Modified Replacement A example. The reduced Sample 2 is appropriate, however, only if all 588 cases in Subclass 1a are reviewed for FS. In Modified Replacement B with $n_{afm} > n_{fma}$ only that portion of Subclass 1a needed to replace Subclass 2a is reviewed for FS, however. Thus, the full-size Sample 2 must be selected to determine Subclasses 2a, 2b, and 2c.

C360 SOCIAL SECURITY NUMBER METHOD.

C361 SAMPLING PARAMETERS.

n_a, n_f, n_m: AFDC, FS, and MAO (non-AFDC) sample sizes.
n'_a, n'_f, n'_m: inflated sample sizes compensating for expected number of not-subject-to-review cases.
N_a, N_f, N_m: estimated average monthly sizes of the frame populations for AFDC, FS, and MAO (non-AFDC).
$R_i = (n'_i)\ (10{,}000) / [(N_i)\ (12)],\ i=a,f,m$.

C362 SUMMARY TABLE[7].

Domain	All cases matching set of R_m #'s	All cases matching set of R_f-R_m #'s	All cases matching set of R_a-R_f #'s
FS/AFDC/Medicaid	1a [FAM]	2a [FAM]	3a [FAM]
AFDC/Medicaid	1b [AM]	2b [AM]	3b [AM]
FS/Medicaid	1c [FM]	2c [F]	3c [Ø]
F.S. only	1d [F]	2d [F]	3d [Ø]
Medicaid only	1e [M]	2e [Ø]	3e [Ø]

7 Table assumes $R_a > R_f > R_m$. The monthly sampling consists of generating R_a number of distinct four-digit random numbers. The resulting numbers are grouped into three sets consisting of the first R_m numbers, the next R_f-R_m numbers, and the last R_a-R_f numbers. Cases are matched to the random four-digit numbers by the last four digits of the social security number for the key person in the sampling unit or, if there is no such identifier, by a four-digit number assigned to the case prior to sampling.

C370 SEQUENTIAL RANDOM SAMPLE METHOD.

C371 SAMPLING PARAMETERS.

n'_a, n'_f, n'_m : sizes of samples drawn from AFDC, FS and Medicaid frames, respectively, determined from precision calculations.

C372 SUMMARY TABLE.

Domain	AFDC cases	Food Stamp cases	Medicaid cases
FS/AFDC/Medicaid	1a [FAM]	2a [FAM]	3a [FAM]
AFDC/Medicaid	1b [AM]		3b [AM]
FS/Medicaid		2b [FM]	3c [FM]
FS only		2c [F]	
Medicaid only			3d [M]

C380 DISPROPORTIONATE STRATIFIED METHOD A.

C381 SAMPLING PARAMETERS.

n_1, n_2, n_3: AFDC, FS, and MAO (non-AFDC) sample sizes.
n_{hi} = Neyman allocation of n_i into stratum h.
G_i = variance for Neyman allocation and "take all strategy" divided by variance for proportional allocation.
$n''_{hi} = G_i \, n'_{hi}$.

C382 SUMMARY TABLE.

Domain	AFDC cases*	FS cases*	Medicaid cases not rcvng AFDC*
FS/AFDC/Medicaid	1.1 (n''_{11}) †	2.1 (n''_{12}) †	
AFDC/Medicaid	1.2 (n''_{21}) [AM]		
FS/Medicaid		2.2 (n''_{22}) §	3.1 (n''_{13})m §
FS only		2.3 (n''_{32}) [F]	
Medicaid only			3.2 (n''_{23}) [M]

*Indicated sample sizes achieved by subsampling from each frame.

†The reviews performed in samples 1.1 and 2.1 are determined by parameters n''_{11} and n''_{12}:

	Sample 1.1	Sample 2.1
$n''_{11} \geq n''_{12}$	[FAM]	[Ø]
$n''_{11} < n''_{12}$	[Ø]	[FAM]

§The reviews performed in samples 2.2 and 3.2 are determined by parameters n''_{22} and n''_{23}:

	Sample 2.2	Sample 3.2
$n''_{22} \geq n''_{23}$	[FM]	[Ø]
$n''_{22} < n''_{23}$	[Ø]	[FM]

C390 MODIFIED DISPROPORTIONATE STRATIFIED METHOD A.

C391 SAMPLING PARAMETERS.

$(X_1, X_2, X_3, X_4, X_5)$ is the solution to a non-linear programming problem.

C392 SUMMARY TABLE[8].

	Strata of Totally Integrated Frame				
Domain	FAM	AM	FM	F only	M only
FS/AFDC/Medicaid	$1(1/X_2)$ [FAM]				
AFDC/Medicaid		$2(1/X_2)$ [AM]			
FS/Medicaid			$3(1/X_3)$ [FM]		
FS only				$4(1/X_4)$ [F]	
Medicaid only					$5(1/X_5)$ [M]

8 Table depicts sampling from a totally integrated frame that identifies all overlap domains. Alternatively, subsampling from separate, overlapping frames can be used to achieve the indicated sample sizes.

STATE-ESTIMATED ERROR RATES

D100 GENERAL. This Exhibit presents formulas for estimating error rates from QC data[1]. You must estimate your negative error rate and report it to us each year (see 3300). The calculations of estimated over- and underpayment error rates and the estimated payment error rate[2] are optional. This Exhibit also presents formulas for calculating sampling variances, standard errors, and confidence intervals. These calculations are also optional, except as required to demonstrate equal-or-better precision for alternative sample designs (see Exhibit E).

Because you select a sample each month, the months, strictly speaking, are separate strata. Ignore this monthly stratification, however, except when there is a between-month change in sampling rates (sampling intervals in systematic sampling). If this occurs, create separate strata that correspond to the different sampling rates and assign months with the same sampling rate to one stratum.

D200 ESTIMATES. A stratified sample in which the stratum sampling rates are all the same is said to be "self weighted". Use the formulas in D210 for (1) unstratified samples or (2) self-weighted stratified samples. Use the formulas in D220 for stratified samples that are *not* self weighted. If you are uncertain whether a stratified sample design is self weighted, use the formulas in D220.

D210 ESTIMATES FOR UNSTRATIFIED DESIGNS.

D211 OVER- AN UNDERPAYMENT ERROR RATES. For unstratified designs, estimate the overpayment (or underpayment) error rate by

r = State's estimate of overpayment (or underpayment) error rate

$$= \bar{x} / \bar{u} \qquad \text{(D1)}$$

where

$$\bar{x} = \sum_i x_i/n_a \ , \qquad \text{(D2)}$$

$$\bar{u} = \sum_i u_i/n_a \ , \qquad \text{(D3)}$$

---

1 In our formulas we use "x" to represent State-determined error amounts and "u" to represent benefit amounts. This is consistent with FNS regulations and FNS Handbook 315.

2 The Hunger Prevention Act of 1988 defines the estimated payment error rate as the sum of the point estimates of the over- and underpayment error rates.

x_i = overpayment (or underpayment) error dollars for the ith completed case in the State's active-case sample (When calculating the overpayment error rate, x_i includes error amounts for both overpayment and ineligible cases and is zero for other types of cases. When calculating the underpayment error rate, x_i includes error amounts for underpayment cases and is zero for all other types of cases.)

u_i = benefit dollars for ith completed case in the State's active-case sample, and

n_a = number of completed active cases in the State sample.

D212 PAYMENT ERROR RATE. Estimate the payment error rate by adding together the estimates of the over- and underpayment error rates.

D213 NEGATIVE ERROR RATE. For unstratified designs, estimate the negative error rate by

p = State's estimate of the negative error rate

$$= n_i/n_n \qquad \text{(D4)}$$

where

n_i = number of invalid negative actions in the State's negative-action sample, and

n_n = number of completed negative actions in the State sample.

D220 ESTIMATES FOR STRATIFIED DESIGNS. There are two different ways to estimate over and underpayment error rates from stratified samples: with a separate ratio estimate, described in D221, or with a combined ratio estimate, described in D222. The separate ratio estimate uses dollar weights, whereas the combined ratio estimate uses caseload weights. The separate ratio estimator is usually more precise than the combined ratio estimator[3]. Use the combined estimator, however, if you know the caseload weights but not the dollar weights. If you use the formulas contained in Exhibit H to estimate the weights, then either estimator gives the same answer.

D221 SEPARATE RATIO ESTIMATES FOR OVER- AND UNDERPAYMENT ERROR RATES. In stratified designs you can estimate the overpayment (or underpayment) error rate by using the following separate ratio estimator:

$$r = \sum_h W'_h r_h \qquad \text{(D5)}$$

where

3 Pages 165-169 of Cochran's Sampling Techniques discuss the statistical properties of the two types of stratified ratio estimators.

W'_h = dollar weight for stratum h; that is, the proportion of all benefit dollars issued to all subject-to-review households that are issued to households in stratum h (See formulas in Exhibit H for estimating dollar weights.); and

r_h = estimated overpayment (or underpayment) error rate for stratum h. (This is calculated from the data and sample size in stratum h using equation (D1).)

D222 COMBINED RATIO ESTIMATES FOR OVER- AND UNDERPAYMENT ERROR RATES. Alternatively, in stratified designs you can use the combined ratio estimator:

$$r = \bar{x}_{st} / \bar{u}_{st} \qquad \text{(D6)}$$

where

$$\bar{x}_{st} = \sum_h W_h \bar{x}_h; \qquad \text{(D7)}$$

$$\bar{u}_{st} = \sum_h W_h \bar{u}_h; \qquad \text{(D8)}$$

W_h = caseload weight for stratum h; that is, the proportion of the total number of subject-to-review cases that belong to stratum h (See formulas in Exhibit H for estimating caseload weights.);

$$\bar{x}_h = (\sum_i x_{ih})/n_{ah}\ ; \qquad \text{(D9)}$$

$$\bar{u}_h = (\sum_i u_{ih})/n_{ah}\ ; \qquad \text{(D10)}$$

x_{ih} = overpayment (or underpayment) error dollars for the ith completed case in stratum h of the State's active-case sample (When calculating the overpayment error rate, x_{ih} includes error amounts for both overpayment and ineligible cases and is zero for other types of cases. When calculating the underpayment error rate, x_{ih} includes error amounts for underpayment cases and is zero for all other types of cases.);

u_{ih} = benefit dollars for ith completed case in stratum h of the State's active-case sample; and

n_{ah} = number of completed active cases in stratum h of the State sample.

D223 PAYMENT ERROR RATE. Estimate the payment error rate by adding together the estimates of the over- and underpayment error rates.

D224 NEGATIVE ERROR RATE. For stratified designs, estimate the negative error rate by

$$p = \sum_h W_h p_h \qquad \text{(D11)}$$

where

W_h = caseload weight for stratum h; that is, the proportion of the total number of subject-to-review cases that belong to stratum h (See formulas in Exhibit H for estimating caseload weights.); and

p_h = estimated negative error rate for stratum h. (This is calculated from the data and sample size in stratum h using equation (D4).)

D300 SAMPLING VARIANCES. The following variance formulas assume simple-random sampling within each stratum. These formulas approximate the variances for systematic sampling if the error cases and error amounts (for active cases) are distributed completely at random within the sampling frame. If this is not the case, and if there is "clumping" of error cases in the frame--that is, the occurrence of an error case at a given position in the sampling frame increases the probability there will be another error case nearby in the sampling frame--then these formulas overestimate the variances for systematic sampling. On the other hand, if there is "dispersal" of error cases within the frame--that is, the occurrence of an error case at a given position in the frame decreases the probability of an error case in a nearby position--then these formulas underestimate the variances for systematic sampling.

If you have a self-weighted stratified sample, you can use the variance estimation formulas for unstratified sampling in D310. This will over-estimate the actual variances when there are differences between the stratum means. If this is unacceptable, use the formulas in D320.

D310 SAMPLING VARIANCES IN UNSTRATIFIED DESIGNS.

D311 VARIANCES OF OVER- AND UNDERPAYMENT ERROR RATES. In an unstratified design use the following formula to estimate the variance of the estimated overpayment (or underpayment) error rate:

$$v(r) = (1-f_a)(s^2_x + r^2 s^2_u - 2rs_{ux})/[n_a(\bar{u})^2] \qquad \text{(D12)}$$

where

$1 - f_a$ = the finite-population-correction factor for the active-case sample (f_a is calculated by dividing n_a by the annual active caseload subject to review.),

$$s^2_x = [\sum_i (x_i - \bar{x})^2]/(n_a-1) = [\sum_i x_i^2 - n_a(\bar{x})^2]/(n_a-1) \ , \qquad (D13)$$

$$s^2_u = [\sum_i (u_i - \bar{u})^2]/(n_a-1) = [\sum_i u_i^2 - n_a(\bar{u})^2]/(n_a-1) \ , \qquad (D14)$$

$$s_{ux} = [\sum_i (u_i - \bar{u})(x_i - \bar{x})]/(n_a-1) = (\sum_i u_i x_i - n_a \bar{u}\,\bar{x})/(n_a-1) \qquad (D15)$$

and r, n_a, x_i, $\bar{x}$, u_i, and $\bar{u}$ are defined in D211.

D312 VARIANCE OF THE PAYMENT ERROR RATE. The sampling variance of the payment error rate can be calculated in one of two ways: from microdata or from available summary statistics. To use the microdata approach, let

$$z_i = x_{1i} + x_{2i} \qquad (D16)$$

where

x_{1i} = overpayment error dollars for the ith completed case in the State's active-case sample (x_{1i} includes error amounts for both overpayment and ineligible cases and is zero for other types of cases.) and

x_{2i} = underpayment error dollars for the ith completed case in the State's active case sample.

Then

$v(r_p)$ = estimated variance of the payment error rate

$$= (1-f_a)(s^2_z + r_p^2 s^2_u - 2r_p s_{uz})/[n_a(\bar{u})^2] \qquad (D17)$$

where

$$s^2_z = [\sum_i (z_i - \bar{z})^2]/(n_a-1) = [\sum_i z_i^2 - n_a(\bar{z})^2]/(n_a-1) \ , \qquad (D18)$$

$$s_{uz} = [\sum_i (u_i - \bar{u})(z_i - \bar{z})]/(n_a-1) = (\sum_i u_i z_i - n_a \bar{u}\,\bar{z})/(n_a-1) \ , (D19)$$

f_a and s^2_u are defined above, and r_p is the estimated payment error rate.

To use summary statistics to calculate $v(r_p)$, use equation (D17) but instead of determining the needed inputs from (D18) and (D19), use the following equations:

$$s^2_z = s^2_{x1} + s^2_{x2} - 2\bar{x}_1\bar{x}_2 n_a/(n_a - 1) \qquad (D20)$$

and

$$s_{uz} = s_{ux1} + s_{ux2} \quad , \tag{D21}$$

where the needed inputs to these equations are calculated from equations (D2), (D13), and (D15) with the subscript of "1" indicating calculations with overissuance errors and the subscript of "2" indicating calculations with underissuance errors.

D313 VARIANCE OF THE NEGATIVE ERROR RATE. The following formula estimates the variance of the estimated negative error rate:

$$v(p) = (1-f_n)p(1-p)/(n_n-1) \tag{D22}$$

where

$1 - f_n =$ the finite-population-correction factor for the negative-action sample (f_n is calculated by dividing n_n by the annual negative caseload subject to review.),

and p and n_n are defined in D213.

D320 SAMPLING VARIANCES FOR STRATIFIED DESIGNS. If you estimate the overpayment (or underpayment) error rate by using the formulas in Exhibit H to estimate stratum weights, then use the equations in D322 to estimate the corresponding sampling variances.

D321 VARIANCES FOR SEPARATE RATIO ESTIMATES.

D321.1 Under- and Overpayment Error Rates. In stratified designs, if you estimate the overpayment (or underpayment) error rate with the separate ratio estimator, the following formula estimates the sampling variance:

$$v(r) = \sum_h (W'_h)^2 v_h(r) \tag{D23}$$

where

$v_h(r) =$ the variance contribution from stratum h, calculated using equation (D12) and only the data, sample size, and annual caseload corresponding to stratum h

and W'_h is the dollar weight for stratum h, defined in D221.

D321.2 Payment Error Rate. In stratified designs, if you use separate ratio estimates for the under- and overpayment error rates, estimate the sampling variance for the estimated payment error rate from the following equation:

$$v(r_p) = \sum_h (W'_h)^2 v_h(r_p) \tag{D24}$$

where

$v_h(r_p)$ = the variance contribution from stratum h calculated from equation (D17) and only the data, sample size, and annual caseload corresponding to stratum h.

D322 VARIANCES OF COMBINED RATIO ESTIMATES.

D322.1 Over- and Underpayment Error Rates. In stratified designs, if you use the combined ratio estimator for the overpayment (or underpayment) error rate, the following formula estimates the sampling variance:

$$v(r) = \{ \sum_h [W_h^2(1-f_{ah})(s^2_{xh} + r^2 s^2_{uh} - 2rs_{uxh})/n_{ah}] \} / \bar{u}_{st}^2 \quad (D25)$$

where r and $\bar{u}_{st}$ are defined in D220; W_h is the caseload for stratum h (also defined in D220); and all the remaining terms are defined in D310, except are subscripted with "h" to indicate that they are calculated only from the data, sample sizes, and annual caseload sizes corresponding to stratum h.

D322.2 Payment Error Rate. In stratified designs, to estimate the variance of the estimated payment error rate when the over-and underpayment error rates are combined ratio estimates, use the following formula:

$$v(r_p) = \{ \sum_h [W_h^2(1-f_{ah})(s^2_{zh} + r_p^2 s^2_{uh} - 2r_p s_{uzh})/n_{ah}] \} / \bar{u}_{st}^2 \quad (D26)$$

where all the terms are the same as in equation (D25) except for r_p, s^2_{zh}, and s_{uzh}. To calculate s^2_{zh} use equation (D18) or (D20) applied to stratum h; for s_{uzh}, use equation (D19) or (D21) applied to stratum h.

D323 VARIANCE OF THE NEGATIVE ERROR RATE. For stratified designs, the following formula estimates the variance of the estimated negative error rate:

$$v(p) = \sum_h (W_h)^2 v_h(p) \quad (D27)$$

where

$v_h(p)$ = the variance contribution from stratum h, calculated using equation (D22) and only the data, sample size, and annual caseload corresponding to stratum h

and W_h is the caseload weight for stratum h, defined in D224.

D400 STANDARD ERRORS. Estimate standard errors by taking the square root of the estimated sampling variances.

D500 CONFIDENCE INTERVALS.

D510 DEFINITION. Your estimated error rate and an estimate of its standard error permit you to calculate a "confidence interval". This is an interval that with prescribed confidence contains the average result of all possible samples of the same design and sample size.

To illustrate, if all possible samples of the same design and sample size were selected, each of the samples were assigned to reviewers to perform QC reviews under essentially the same conditions, and an estimated error rate and its estimated standard error were calculated from each sample then:

A Approximately 9/10 of the intervals from 1.64 standard errors below the estimate to 1.64 standard errors above the estimate would include the average value of all possible samples. An interval from 1.64 standard errors below the estimate to 1.64 standard errors above the estimate is called a 90 percent confidence interval.

B Approximately 19/20 of the intervals from 1.96 standard errors below the estimate to 1.96 standard errors above the estimate would include the average value of all possible samples. An interval from 1.96 standard errors below the estimate to 1.96 standard errors above the estimate is called a 95 percent interval.

C Almost all intervals from three standard errors below the estimate to three standard errors above the estimate would include the average value of all possible samples.

The average value of all possible samples of the same design and sample design may or may not be contained in any particular computed interval. But for a particular sample, you can say with specified confidence that the average of all possible samples in included in your calculated confidence interval.

D520 EXAMPLE. For example if the estimated error rate is 5.17 percent and the standard error is 1.13 percent then the 90-percent confidence interval is from

$$5.17 - (1.64)\ (1.13) = 3.32 \text{ percent}$$

to

$$5.17 + (1.64)\ (1.13) = 7.02 \text{ percent.}$$

COMPARING PRECISIONS OF STRATIFIED AND UNSTRATIFIED SAMPLE DESIGNS

E100 GENERAL. 7 CFR 275.11(b)(4) establishes the equal-or-better precision requirement for alternative sample designs. It states that an alternative sample design must provide equal or better precision than the smallest-permitted simple-random sample. For the active sample, the precision of interest is the precison of the payment error rate[1]; and for the negative sample, that of the negative error rate.

The equal-or-better-precision requirement applies to the planning of your sampling activities and not to the actual statistical performance of your sample. Thus, the precisions of concern are predicted precisions and not end-of-year calculated precisions.

This Exhbit explains how to compute the predicted precisions needed to demonstrate equal-or-better precision for stratified samples. Use the procedures described in this Exhibit if you find yourself in any of the following situations:

A New Stratified Sample Design. If your present sample for the current year is unstratified or has different strata definitions from your proposed stratified sample for the coming year, you must demonstrate that the proposed stratification satisfies the equal-or-better-precision requirement[2].

B Unchanged Stratified Sample Design. If your sample for the current year is disproportionately stratified and you propose to use the same sample design for the coming year, you must demonstrate with current data that the equal-or-better-precision requirement continues to be satisfied. (For the active-case sample, if previous demonstrations compared precisions of the case error rate or the overpayment error rates, the updated demonstration of equal-or-better precision must compare the precisions for the payment error rate.)

C Stratification by Time in Order To Adjust the Sample Size. If you propose in the middle of the year to correct for under- or oversampling by changing the sampling rate, then you must demonstrate that the resulting stratified design satisfies the equal-or-better precision design. The calculations for this situation are much simpler than for the preceding situations—see E340.

---

1 The Hunger Prevention Act of 1988 defines the estimated payment error rate as the sum of the point estimates of the estimated over- and underpayment error rates.

2 You do not have to demonstrate equal-or-better precision if you are proposing proportional allocation of a sample of size greater than or equal to your minimum sample size (as specified in 3121.1).

E200 STRATIFIED SAMPLES THAT ARE PROPORTIONALLY ALLOCATED. A stratified sample that is proportionately allocated has equal or better precision than a simple random sample of the same size. Therefore, if you are proposing proportional allocation of your minimum sample size or of a larger sample size, you need only to document in your sample plan the mechanics of your sample allocation. If you are proposing, however, to proportionately allocate a sample size smaller than your minimum sample size, you must calculate predicted variances using the procedure described in E300.

E300 STRATIFIED SAMPLES THAT ARE DISPROPORTIONATELY ALLOCATED. If you are proposing a disproportionately stratified sample, you must compare the predicted sampling variances of the proposed stratified sample with that of the smallest-allowed simple-random sample. You can make this comparison by determining if the ratio of the predicted variances is greater than or less than 1.0. Since dividing both the numerator and denominator of a ratio by the same number leaves the ratio unchanged, you can ignore any multiplicative terms that appear in both predicted variances. This greatly simplifies the calculations for those samples in which the only stratification is by time. E340 describes these simpler calculations.

When a sample is stratified on some basis other than by time, we say it is non-temporally stratified. E310 and E320 describe for the active and negative samples, respectively, the precision calculations for non-temporal stratification.

E310 NON-TEMPORAL STRATIFICATION OF THE ACTIVE-CASE SAMPLE. Use the following procedure to demonstrate equal-or-better precision for a non-temporally stratified sample of active cases. This procedure parallels the six steps of calculations described on pages 16 and 17 of the AFDC sampling manual[3]. FNS regional statisticians can assist you with these calculations by (1) estimating from your transmitted QC data the parameters described in Step 1, below, and (2) using an FNS-developed SuperCalc4 spreadsheet, called ACCOMPARE, to do the actual computations based on your minimum sample size and proposed sample allocation.

E311 STEP 1: ESTIMATE STRATUM- AND POPULATION-LEVEL PARAMETERS. You need a recent data set of active cases. It must contain the error and benefit amounts for all your completed cases for a period of at least six months. It can be stratified or unstratified. If stratified, it need not be stratified in the same way as your proposed sample. It must also contain stratum identifiers—both for the proposed stratification and for any stratification in effect at the time of the data set's

3 U.S. Department of Health and Human Services, Office of Family Assistance. Revised AFDC Quality Control Manual, Section 2, Sampling and Statistical Methods. January 17, 1985.

sample selection. You will use this data set to estimate needed stratum- and population-level parameters.

You must estimate six parameters for each proposed stratum and for the entire population. To each parameter there corresponds a subset of your data set that we call corresponding data. Specifically, the entire data set corresponds to the population parameters, and the cases in proposed stratum h correspond to the parameters needed for stratum h. Estimate each needed parameter with the appropriate equation listed below. The choice of the appropriate equation depends of whether the corresponding data are (1) unstratified or proportionally stratified or (2) disproportionately stratified.

			equation	
name	symbol for population	symbol for stratum h	unstratified or proportionally stratified	disproportionately stratified
Payment error rate	r_p	r_{ph}	See D212[4]	See D223[4]
Average error dollars	$\bar{z}$	$\bar{z}_h$	(E1)[4]	(E1)[4]
Average benefit dollars	$\bar{u}$	$\bar{u}_h$	(D3)	(D8)
Variance of the error dollars	s^2_z	s^2_{zh}	(D18) or (D20)	(E2)
Variance of the benefit dollars	s^2_u	s^2_{uh}	(D14)	(E3)
Covariance between error dollars and benefit dollars	s_{uz}	s_{uzh}	(D15)	(E4)

Equations (E1) through (E4) appear below. Each of these equations can be applied to your entire data set (when estimating population parameters) or just to the cases in proposed stratum h (when estimating stratum parameters).

$$\bar{z} = r_p\bar{u} \ , \qquad \text{(E1)}$$

where r_p is the payment error rate (see D212 and D223) for the

4 Alternatively, you can calculate $\bar{z}$ as the average over the corresponding data of the microdata defined by equation (D16). (If the corresponding data is disproportionately stratified, calculate a (caseload) weighted average.) Then

$$r_p = \bar{z}/\bar{u}.$$

corresponding data and $\bar{u}$ is the average benefit amount (see equations (D3) and (D8)) for the corresponding data.

Equations (E2) through (E4), below, follow from pages 136 and 137 of Cochran's Sampling Techniques (third edition). These equations assume that the sample size is large. The subscript of "k" indexes the strata that were in effect when the cases in your data set were selected.

$$s^2_z = \sum_k W_k \left(\sum_i z_{ik}^2\right)/n_{ak} - (\bar{z})^2 \quad , \tag{E2}$$

$$s^2_u = \sum_k W_k \left(\sum_i u_{ik}^2\right)/n_{ak} - (\bar{u})^2 \quad , \tag{E3}$$

$$s_{uz} = \sum_k W_k \left(\sum_i u_{ik} z_{ik}\right)/n_{ak} - \bar{u}\,\bar{z} , \tag{E4}$$

where W_k is the $k\underline{th}$ stratum's caseload weight[5] with respect to the corresponding data and n_{ak} is the number of completed cases in the $k\underline{th}$ stratum. $\bar{z}$ and $\bar{u}$ in equations (E2) through (E4) are (caseload) weighted averages over the corresponding data.

EXAMPLE. Consider the following data set[6] and proposed strata:

h	k	N_k	u_{ik}	z_{ik}
10	1	10,000	100	0
			150	0
			30	30
	2	15,000	150	0
			75	10
20	3	25,000	40	0
			50	5

Then

$f_1 = 3/10000 = .00030$
$f_2 = 2/15000 = .00013$
$f_3 = 2/25000 = .00008$.

5 Caseload weights are with respect to subject-to-review cases (see D222).

6 This example, which we provide only for purposes of illustrating the necessary calculations, does not satisfy the assumption (for equations (E2) through (E4)) that the sample size is large.

Thus, neither the entire data set nor proposed stratum 10 is proportionally allocated. Calculating the caseload weights gives:

$$W_1 = 10000/50000 = .2$$
$$W_2 = 15000/50000 = .3$$
$$W_3 = 25000/50000 = .5$$

$$W_1/(W_1+W_2) = .4$$
$$W_2/(W_1+W_2) = .6 \ .$$

The estimated parameters are

$$\bar{z} = (.2)(10) + (.3)(5) + (.5)(2.5) = 4.75$$
$$\bar{u} = (.2)(93.3) + (.3)(112.5) + (.5)(45) = 74.91$$

$$r_p = 4.75/74.91 = .0634$$
$$s^2_z = (.2)(300) + (.3)(50) + (.5)(12.5) - (4.75)^2 = 58.7$$
$$s^2_u = (.2)(11133) + (.3)(14062.5) + (.5)(2050) - (74.91)^2 = 1858.8$$
$$s_{uz} = (.2)(300) + (.3)(375) + (.5)(125) - (4.75)(74.91) = -120.8$$
$$\bar{z}_{10} = (.4)(10) + (.6)(5) = 7.00$$
$$\bar{u}_{10} = (.4)(93.3) + (.6)(112.5) = 104.82$$

$$r_{p10} = 7.00/104.82 = .0668$$
$$s^2_{z10} = (.4)(300) + (.6)(50) - (7.00)^2 = 101$$
$$s^2_{u10} = (.4)(11133) + (.6)(14062.5) - (104.82)^2 = 1903.5$$
$$s_{uz10} = (.4)(300) + (.6)(375) - (7.00)(104.82) = -388.74$$
$$\bar{z}_{20} = 2.5$$
$$\bar{u}_{20} = 45$$

$$r_{p20} = 2.5/45 = .0556$$
$$s^2_{z20} = (2.5^2 + 2.5^2)/1 = 12.5$$
$$s^2_{u20} = (5^2 + 5^2)/1 = 50.0$$
$$s_{uz20} = [(-5)(-2.5) + (5)(2.5)]/1 = 25 \ .$$

E312 STEP 2: PREDICT UNSTRATIFIED VARIANCE. Use equation (D17) to predict the unstratified sampling variance, v_{ran}. $\bar{u}$, r_p, s_{2z}, and s^2_u are the estimated population parameters from Step 1. n_a is your minimum active-case sample size for simple-random sampling, which will probably be different from the number of cases in the data set you used in Step 1.

E313 STEPS 3 & 4: PREDICT INDIVIDUAL STRATA VARIANCES. Use equation (D17) to predict the sampling variance for each proposed stratum. $\bar{u}$, r_p, s^2_z, and s^2_u are the estimated stratum parameters from Step 1. n_a is the number of reviewable cases that you propose to allocate to the particular stratum.

E314 STEP 5: PREDICT STRATIFIED VARIANCE. Use equation (D24) to predict the stratified sampling variance, v_{strat}. $v_h(r_p)$ is the result for stratum h from Steps 3 and 4.

E315 STEP 6: COMPUTE VARIANCE RATIO.

$$\text{RATIO} = v_{strat}/v_{ran} \cdot$$

If RATIO is less than or equal to 1.0, your proposed sample design satisfies the equal-or-better precision requirement.

Since RATIO depends on estimated parameters from Step 1, the calculated value of RATIO is also an estimate. Because of sampling variability, the calculated value of RATIO may be greater than 1.0 even if the actual value of RATIO is less than or equal to 1.0. Consequently, the (FNS-developed) ACOMPARE spreadsheet calculates 95-percent, 87-percent, and 75-percent lower confidence limits for RATIO based on the its estimated standard error.

We will conclude that your sample design fails the equal-or-better-precision requirement if the following criterion is satisfied:

A The 95-percent lower confidence limit exceeds 1.0, OR

B The 87-percent lower confidence limits exceed 1.0 for any two out of three years, OR

C The 75-percent lower confidence limits exceed 1.0 for any three out of four years, OR

D The calculated values of RATIO exceed 1.0 for five consecutive years.

If the true value of RATIO is really less than or equal to 1.0, the probability this criterion will be satisfied is less than 0.05—that is, a chance of less than one in twenty.

E320 NON-TEMPORAL STRATIFICATION OF THE NEGATIVE-ACTION SAMPLE. Use the following procedure to demonstrate equal-or-better precision for a non-temporally stratified sample of negative actions. FNS regional statisticians can assist you with these calculations by using an FNS-developed SuperCalc4 spreadsheet, called NCOMPARE, to do the actual computations based on (1) the negative error rates you estimate in Step 1, below, (2) your minimum sample size, and (3) your proposed sample allocation.

E321 STEP 1: ESTIMATE STRATUM- AND POPULATION-LEVEL PARAMETERS. You need a data set of at least six months of sampled negative actions in order to estimate the negative error rate for the entire population and for each proposed stratum. This data set must contain the error findings and stratum identifiers--both for the proposed stratification and any stratification in effect at the time of the data set's sample selection. Use equation (D4) to estimate a needed negative error rate if the corresponding data are unstratified or proportionately stratified. Use equation (D11) if the corresponding data are disproportionately stratified.

E322 STEP 2: PREDICT UNSTRATIFIED VARIANCE. Use equation (D22) to predict the unstratified sampling variance, v_{ran}. p is the estimated population negative error rate from Step 1. n_h is your minimum negative-action sample size for simple-random sampling, which will probably be different from the number of cases in the data set you used in Step 1.

E323 STEPS 3 & 4: PREDICT INDIVIDUAL STRATA VARIANCES. Use equation (D22) to predict the sampling variance for each proposed stratum. p is each proposed stratum's negative error rate from Step 1. n_h is the number of reviewable cases that you propose to allocate to the particular stratum.

E324 STEP 5: PREDICT STRATIFIED VARIANCE. Use equation (D27) to predict the stratified sampling variance, v_{strat}. $v_h(p)$ is the result for stratum h from Steps 3 and 4.

E325 STEP 6: COMPUTE VARIANCE RATIO. This step for the negative sample is the same as for the active sample (see E315), except that the (FNS developed) spreadsheet NCOMPARE provides lower confidence limits for RATIO.

E330 END-OF-YEAR MINIMUM SAMPLE SIZES FOR ALTERNATIVE SAMPLE DESIGNS. At the end of the year, the minimum sample size for an alternative sample design equals the sample size specified in the sample plan (or in conjunction with a mid-year sampling correction) multiplied by the following:

A If RATIO is greater than 1.0 (but we did not conclude that the sample design failed the equal-or-better-precision requirement), then the multiplying factor, called NFACTOR, only adjusts for the difference between the actual caseload and the estimated caseload as follows:

1 If the actual caseload is larger than the estimated caseload but by less than 20 percent, then NFACTOR = 1.0.

2 If the actual caseload is larger than the estimated caseload by more than 20 percent or is less than the estimated caseload, then

$$\text{NFACTOR} = n'_{min}/n_{min} ,$$

where

n'_{min} = minimum sample size for simple random sampling based on your actual caseload, and

n_{min} = minimum sample size for simple random sampling based on your estimated caseload .

B If RATIO is less than 1.0, then the multiplying factor is (RATIO) (NFACTOR) .

E340 STRATIFICATION BY TIME. The comparison of precisions simplifies considerably if the only stratification is by time. The reason is our analysis of QC data indicates that population means, variance and covariances are approximately constant across months. Thus, these terms can be cancelled out of both the numerator and denominator of RATIO (see E315 and E316) when there is no other type of stratification. In this situation, RATIO will be less than or equal to 1.0 (indicating that the sample design satisfies the equal-or-better-precision requirement) if

$$\sum_{h=1}^{L} W_h^2/n_h \leq 1/n_{min}$$

where

W_h = the caseload weight for stratum h[7],

n_h = number of reviewable cases in stratum h, h =1, 2, . . . , L, and

n_{min} = the minimum sample size for simple-random sampling.

Let

$$A = 1/n_{min} - \sum_{h=1}^{L-1} W_h^2/n_h \ .$$

If A is zero or negative, then the stratified design fails to satisfy the equal-or-better-precision requirement. If A is positive, then

$$n_L \geq W_L^2/A$$

assures that the stratified design satisfies the equal-or-better-precision requirement.

7 Caseload weights are defined with respect to subject-to-review cases (see D222 and D224).

Special Case. If there are only two strata, then the stratified design satisfies the equal-or-better-precision requirement if

$$n_1 - W_1^2 n_{min} > 0$$

and

$$n_2 \geq (n_{min} - n_1) + (n_1 - W_1 n_{min})^2 / (n_1 - W_1^2 n_{min}) \ .$$

E341 EXAMPLE. At the beginning of the year, you predict the average-monthly subject-to-review caseload will be 45,000. You have signed a reliability waiver, so n_{min} = 930 and n_{min}/12 = 77.5. You sample systematically. Four months into the year, however, the following are the sampling parameters by month:

N	n (reviewable)
30,000	52
25,000	43
26,000	45
24,000	41
105,000	181

where N is each month's subject-to-review caseload calculated from one of reviewable-caseload the equations in 3320; and n is either (1) an exact count of each month's reviewable cases (if all cases in the month have been disposed) or (2) an estimate obtained by adjusting the number of selected cases downward by the anticipated number of Not Subject To Review cases.

In light of this information, you predict that the average monthly caseload for the remaining eight months will be 25,000. Thus, the updated, subject-to-review

$$\underline{\text{annual-average}} \text{ monthly caseload} = [105{,}000 + (8)(25{,}000)] \ / \ 12$$
$$= 305{,}000 \ / \ 12 = 25{,}417 \ ,$$

which implies n_{min} = 578. If you continue using the same sampling interval, your expected sample size at the end of the year will be

$$181 + (8)(77.5)(25{,}000)/45{,}000 = 525.$$

Thus, you must correct for undersampling, which you propose to do by changing your sampling interval. The resulting stratified design has the following caseload weights:

$$W_1 = 105{,}000 \ / \ 305{,}000 = .344 \qquad W_2 = 1 - .344 = .656 \ .$$

The number of reviewable cases that you need from the remaining eight months, denoted n_2, in order to satisfy the equal-or-better precision requirement is calculated as follows:

$$n_{min} - W_1{}^2 n_{min} = 181 - (.344)_2(578) = 112.6$$

$$n_2 \geq 578 - 181 + [181 - (.344)(578)]^2/112.6 = 397 + 2.8 = 400$$

or

$$n_2/8 \geq 50.$$

Assume you change your sampling interval to provide for the remaining eight months the needed 400 reviewable cases. After another four months, however, the resulting sampling parameters are:

N	n (reviewable)
21,000	42
20,000	40
19,000	38
20,000	40
80,000	160

In light of this additional information, you predict that the average monthly caseload for the the last four months will be 19,000. Thus, the updated, subject-to-review

$$\underline{\text{annual-average}}\text{ monthly caseload} = [105{,}000 + 80{,}000 + (4)(19{,}000)] / 12 = 261{,}000 / 12 = 21{,}750,$$

which implies $n_{min} = 512$. If you continue using the same sampling interval, your expected sample size at the end of the year will be

$$181 + 160 + (4)(50)(19{,}000)/25{,}000 = 493 .$$

This is less than

$$n_{min}' = (181 + 400)(512)/578 = 515 ,$$

which is the projected minimum sample size at the end of the year for the two-stratum sample design. Hence, another correction for undersampling is needed. The following calculates the number of reviewable cases needed for the last four months:

$$W_1 = 105{,}000/261{,}000 = .402$$
$$W_1 = 80{,}000/261{,}000 = .307$$
$$W_3 = 1 - .402 - .307 = .291$$

$$A = 1/512 - (.402)^2/181 - (.307)^2/160 = .0004713$$

$$n_3 \geq (.291)^2/.00004713 = 180 .$$

Then

n_{min}'' = projected minimum sample size at the end of the year for the three-stratum sample design

$= n_1 + n_2 + n_3$

$= 181 + 160 + 180 = 521$

[Note that $n_{min}'' = 521 > n_{min}' = 515 > n_{min} = 512$.]

ILLUSTRATION OF THE PROPERTIES OF SAMPLE-BASED ESTIMATORS

Consider a population consisting of six elements with values of 1, 2, 3, 6, 7, and 8. Then the population mean is

$$(1+2+3+6+7+8)/6 = 27/6 = 4.5.$$

Column 1 of the following table lists the 15 different ways to select simple random samples of size 4 from this population.

sample	sample mean	squared difference
3 6 7 8	6.00	2.25
2 6 7 8	5.75	1.56
2 3 7 8	5.00	.25
2 3 6 8	4.75	.06
2 3 6 7	4.50	.00
1 6 7 8	5.50	1.00
1 3 7 8	4.75	.06
1 3 6 8	4.50	.00
1 3 6 7	4.25	.06
1 2 7 8	4.50	.00
1 2 6 8	4.25	.06
1 2 6 7	4.00	.25
1 2 3 8	3.50	1.00
1 2 3 7	3.25	1.56
1 2 3 6	3.00	2.25
TOTAL :	67.5	10.38
TOTAL/15 :	4.5 (Expected Value)	.69 (Variance)

Column 2 shows the sample mean associated with each sample. Each of these sample means is an estimate of the population mean of 4.5. Since the average over all 15 samples of the sample mean is 4.5, the sample mean is an unbiased estimator of the population mean. Column 3 shows the squared difference between each sample mean and 4.5. The average squared difference—that is, the variance of the estimator—is 0.69[1]. Hence, the standard error is

$$\sqrt{0.69} = 0.83 .$$

1 This illustrates the calculation of the sampling variance as it is defined in 2310; that is, calculated from the estimates associated with all possible samples. In applications, however, the sampling variance can usually be estimated from the data association with a single sample. For example, see Exhibit D.

We now define the three elements whose values are 1, 2, and 3 to be stratum 1 and the three elements whose values are 6, 7, and 8 to be stratum 2. Column 1 of the next table lists the nine different ways to select a proportionally stratified sample of size 4 from the population.

sample						sample mean	squared difference
	2	3		7	8	5.00	.25
	2	3	6		8	4.75	.06
	2	3	6	7		4.50	.00
1		3		7	8	4.75	.06
1		3	6		8	4.50	.00
1		3	6	7		4.25	.06
1	2			7	8	4.50	.00
1	2		6		8	4.25	.06
1	2		6	7		4.00	.25
TOTAL :						40.5	.75
TOTAL/9 :						4.5 (Expected Value)	.083 (Variance)

The average of the sample means over the nine samples is 4.5. Thus, the sample mean is again an unbiased estimator of the population mean. The standard error is

$$\sqrt{.083} = 0.29 \ .$$

Since the variance of the proportionally stratified sample (0.083) is smaller than the variance of the simple random sample (0.69), the proportionally stratified sample has better precision than the random sample.

EXAMPLE OF A STATE RELIABILITY WAIVER STATEMENT

I (We), ____,
State Official(s) Name(s)

the ____,
Position(s)

certify that I/we have the authority to enter into binding agreements on behalf of the ____________________ Food Stamp Program.
State

Pursuant to 7 CFR 275.11(a)(2)(viii), I/we, acting in my/our official capacity(ies), elect on behalf of the State to exercise the option to reduce the standard Food Stamp-QC active sample by determining the minimum sample size on the basis of 7 CFR 275.11(b)(1)(iii) instead of 7 CFR 275.11(b)(1)(ii). Based on an estimated average monthly caseload of _________ subject-to-review cases, this election reduces our minimum sample size of _________ cases to _________ cases for the annual sample period, October 1, 19 ___ through September 30, 19 ___.

If at the end of this annual sample period the <u>actual</u> average monthly caseload of subject-to-review cases differs from the <u>estimated</u> average monthly caseload, I/we agree that the value of "n" used in the formula specified in 7 CFR 275.11(b)(1)(iii) will be determined according to FNS Handbook 311, Section 3121.1 (B).

I/we recognize that by electing this option the Food Stamp error rate based on this reduced sample size may have less precision than the error rate produced by the standard sample size. Because of this possible effect, the State waives the right to challenge the precision of the resulting Food Stamp error rate based on the reduced sample size.

____________________	____________________
Signature	Signature
____________________	____________________
Title	Title
____________________	____________________
Date	Date

ESTIMATION OF STRATUM WEIGHTS FOR STRATIFIED SYSTEMATIC SAMPLES[1]

H100 GENERAL. Exhibit D defines the two kinds of stratum weights—caseload weights and dollar weights—as follows:

W_h = caseload weight for stratum h

= the proportion of the total number of subject-to-review cases that belong to stratum h

and

W'_h = dollar weight for stratum h

= the proportion of all benefit dollars issued to subject-to-review households that are issued to households in stratum h.

Because these weights pertain to subject-to-review cases, you may be uncertain of the correct values of your stratum weights when your sampling frames contain many listed-in-error cases. Consequently, this Exhibit presents formulas to estimate stratum weights in stratified systmatic samples. Though these formulas can also be applied to some integrated sample designs (that employ systematic sampling), we confine the following discussion to non-integrated samples.

H200 CASELOAD WEIGHTS. The formula for estimating the caseload weight is

$$\hat{W}_h = \hat{N}_h(\text{population}) \ / \ \hat{N}(\text{population}) \qquad \text{(H-1)}$$

where

$$\hat{N}_h(\text{population}) = [n_h(\text{select}) - n_h(\text{NSR})] \ I_h; \qquad \text{(H-2)}$$

1 The weights calculated here sum to 1.0 when added across strata. This is not a necessary requirement for weights used in a combined ratio estimate (see D222). For example, if the weights were to sum to 2.0, then the factor of 2.0 would cancel out from the numerator and denominator of the combined ratio estimate. For the separate ratio estimate, however, it is necessary for the weights to sum to 1.0 (see D221).

n_h(select) = number of cases selected from the sampling frame during the year in stratum h (Include those selected cases that are subsequently deselected in a correction for oversampling. Also, include cases selected by supplemental sampling (either from the full sampling frame or from a reserve pool) to correct for undersampling. Exclude those cases placed into a reserve pool but never removed from the pool to provide additional reviews to correct for undersampling.);

n_h(NSR) = number of selected cases disposed of as not subject to review during the year in stratum h (Exclude those cases placed into a reserve pool but never removed from the pool to provide additional reviews to correct for undersampling.);

I_h = the effective sampling interval (see 3122.2) for stratum h (If during the year you adjusted the sample in stratum h to correct for under- or oversampling, then I_h is the revised effective sampling interval. If you corrected for undersampling by reviewing additional cases from a reserve pool but did not calculate a revised interval, then you must adjust I_h before using it in equation (H-2)—see H400.); and

$\hat{N} = \sum \hat{N}_h$.

H300 DOLLAR WEIGHTS. The formula for estimating the dollar weight is

$$\hat{W}'_h = \hat{N}_h\bar{u}_h \, / \, \sum_k \hat{N}_k\bar{u}_k \qquad \text{(H-3)}$$

where $\hat{N}_h$ is defined in H200 and $\bar{u}_h$ is defined in D220.

H400 SPECIAL SITUATION. If you corrected for undersampling by reviewing additional cases from a reserve pool but did not calculate a revised interval, then you must adjust I_h before using it in equation (H-2). If you use the method described in 3122.3 to select cases for the reserve pool, then use the following formula to adjust I_h for its use in equation (H-2):

$$I'_h = \text{adjusted value of } I_h \qquad \text{(H-4)}$$

$$= I_{ah}I_{ph} \, / \, (I_{ph} - 1 - f_h)$$

where

I_{ah} = the actual sampling interval in stratum h for selecting cases (both review and reserve) from the sampling frame,

I_{ph} = the subsampling interval in stratum h for selecting reserve cases from the set of sampled cases,

f_h = n_{wh}/n_{ph},

n_{wh} = number of cases withdrawn from the reserve pool in stratum h and added to the set of cases to be reviewed, and

ph = total number of cases put into the reserve pool for stratum h.

Note that if f_h=1, then $I'_h = I_{ah}$.

SOURCES OF RANDOM NUMBERS

I100 COMPUTER GENERATION OF RANDOM NUMBERS.

I110 SYSTEM-CLOCK METHOD. An ad-hoc method for producing a four-digit (or less) random number is to scale the system-clock values of unit-seconds and milliseconds. For example, if the system clock reads 22:15:37.175, you would multiply 0.7175 by the sampling interval (assumed to be an integer between 1 and 9999), add 1.0, and round down to the nearest integer. This method produces one random number that is uniformly distributed over the range of the sampling interval. The drawback of this method is that if it is used in a program to generate two or more random numbers, the system clock will change very little during the time the random numbers are generated. Thus, the resulting random numbers will not be randomly distributed but will instead be serially correlated. Reversing the digits (thus, 0.7175 becomes 0.5717) reduces serial correlation between resulting numbers but does not eliminate it. Serial correlation can also be reduced by producing only one random number for each execution of the computer program and then re-executing the program to obtain additional random numbers. Another drawback of this method is that it is sensitive to the computer operator's start-up procedures. For example, if your program is one of the first programs to execute following start-up, then a random number based on the system clock will over-represent low numbers and under-represent high numbers. The reason for this, is that the system clock is usually started by the operator typing into the console terminal the time of day to the nearest whole minute. Related to this phenomenon, is the effect of deferred-execution directives to the operating system. For example, if you direct the operating system to execute your job at 5:00, then the time at which your program reads the system clock will not be a random event. Because of the various drawbacks of this method, we recommend it only if your program is generating only one random number and its execution environment assures that the reading of the system clock is a random event.

I120 PSEUDO-RANDOM NUMBER GENERATORS. A better method for generating random numbers is to use special computer programs, called pseudo-random number generators. These are programs that have been tested and refined to produce numbers that satisfy various statistical tests for randomness. Pseudo-random number generators are sometimes present in a programming language's function library. Other sources include subroutine libraries available from computer users groups or commercial software companies or accompanying textbooks on numerical methods for computers. An example of the latter is the book by Kahener, Moler, and Nash (1989), which contains a 5.25 inch floppy disk of FORTRAN subroutines[1]. Unfortunately, not all pseudo-random number generators included in commercial packages have been sufficiently tested.

1 Kahener, David; Cleve Moler; and Stephen Nash (1989). Numerical Methods and Software. Prentice Hall.

Therefore, you should inquire about the statistical tests a pseudo-random number generator has been subjected to before you commit to using it. If you do use a pseudo-random number generator, be sure that it is designed to produce uniformly distributed random numbers.

Most pseudo-random generators require a "seed value" to initialize their calculations. Therefore, to get different sets of random numbers from separate runs of your program, you must provide different seed values for each run. Some possibilities for seed values are the following:

A The date, if you execute your program on different days.

B Values based on the system clock, providing that the time of execution is a random event and not controlled by system start-up or deferred execution directives (see I110).

C Numbers obtained from a table of random digits (see I200).

D An arbitrary number as seed value for the first execution, followed by an outputted random number from the immediately previous run being used as the seed value in succeeding runs.

I200 TABLE OF RANDOM DIGITS. The digits in a table of random digits appear to be determined entirely by chance. You obtain a sequence of random numbers from such a table by grouping together a fixed number of digits as you progress through the table. Since the individual digits are entirely random, you obtain a sequence of random numbers from any combination and sequence of digits, obtained from any progression (in any direction) through the table, starting from any randomly selected starting point.

For simplicity of use, published tables of random digits often appear in the form of separate columns of five-digit numbers. Both rows and columns may be consecutively numbered for easy reference. The fixed number of digits to be used for a given sequence of random numbers depends on the value of the largest desired random number. For example, if the population to be sampled consists of 84 cases, numbered 1 through 84, random numbers of two digits are required. Similarly, if in another population the highest number assigned is 796, random numbers of three digits are required. To obtain a two-digit,three-digit, seven-digit, or other size number from the table, combined adjacent digits as needed. It makes no difference where in the table you begin or in which direction you move in selecting random numbers as long as you start in a different place each time you use the table.

Example: If the highest consecutively numbered case in the population is 7,543, assume that a randomly selected location starts with the five digits in line 129 of column 4 of Table I1. Assume also that it is decided in advance that the numbers to be used in drawing the sample will be consecutive numbers obtained by reading across the columns from left to right on each consecutive line in the table until a sample of the desired size has been accumulated. If the first four digits of each number in each five-digit column are used, the sample would consist of cases identified as 4140, 1881, 5869, 2925, 5548, 433, 43, 8317, 7015, 8933, 9948, 2446, and so on until the desired sample size is obtained. The numbers 8317, 8933, 9948, as well as any other number larger than 7543 that may later be encountered are not usable for this universe and are, therefore, rejected.

Table I1 lists 10,000 random digits. We generated this table from a FORTRAN program that called the UNI subroutine from the book by Kahaner, Moler, and Nash (1989).

Table I1. 10,000 Random Digits

	1	2	3	4	5	6	7	8	9	10	
001	77553	84790	29740	01754	12167	52710	88928	18235	38745	56447	001
002	27817	68856	83913	57706	83064	53457	21739	60178	39246	04462	002
003	51706	39460	87187	89961	99675	18564	94136	63402	47289	98257	003
004	98784	94193	40943	70962	04056	28220	43650	89768	93126	53382	004
005	77616	17523	35918	53754	08389	13665	70349	98104	21301	59039	005
006	29405	51439	93471	31327	85330	63946	47791	75617	23290	33306	006
007	18149	34562	57636	12270	75633	59609	26558	19474	32497	56789	007
008	56293	88971	20293	36739	85895	79689	02649	01794	47385	66925	008
009	73389	59414	89870	56851	75850	28032	11922	26594	37914	95861	009
010	05748	16637	41382	98646	85467	66794	21068	48411	22796	12053	010
011	43331	53679	24141	79271	50308	75974	69195	54244	55627	54958	011
012	32180	71403	13304	11818	85454	28240	23897	58898	16355	58804	012
013	16771	37774	74063	55451	43673	32713	56664	43220	46679	45981	013
014	01079	96387	98240	36062	17384	65367	58131	59868	88672	89262	014
015	68291	09372	42298	83068	14657	84711	88555	56350	56512	72399	015
016	85791	08892	34703	72929	04402	65438	88550	87391	35870	82946	016
017	08303	62404	79240	80228	46570	66999	74408	16587	84323	89416	017
018	52836	22523	87004	88003	27375	58057	81069	87568	86294	09382	018
019	25311	14264	73817	28124	92838	90269	52031	01522	62525	07916	019
020	57107	80339	07154	14786	21518	16676	62155	96893	23837	22281	020
021	27054	91912	00228	80670	77340	79194	05349	91847	80262	56029	021
022	07432	33631	87914	90602	95195	16360	34228	30342	38136	87482	022
023	92381	16807	68170	32155	14931	44069	41034	86696	94535	11665	023
024	30838	84089	05216	60869	85475	19719	61437	90834	05230	77610	024
025	52975	39655	85020	13212	32118	10215	86288	05720	76402	24246	025
026	47425	41582	02017	18677	45852	82693	06540	67092	82736	46549	026
027	83524	24206	87374	93427	94669	73999	26785	32419	66146	70962	027
028	53916	73928	50758	86501	87861	66174	55598	61367	11929	19361	028
029	92723	50166	69349	17197	61023	67011	60008	69401	87189	65816	029
030	41194	95181	09724	26983	99749	20556	19624	24066	09873	01915	030
031	74578	64960	12129	64942	27917	18665	52534	26366	73163	70354	031
032	42489	27009	19607	17501	70376	29672	82816	08505	46615	86533	032
033	25674	81128	91810	33192	12730	66009	53370	60446	49206	07890	033
034	98129	09267	07110	17117	29745	23240	97643	84564	75635	19312	034
035	12823	63580	75562	17475	26497	67370	20920	44800	59648	94092	035
036	08849	13552	28789	33968	32068	54005	38603	62500	53702	13206	036
037	78806	48854	11454	19424	29423	24294	34314	32095	74451	25312	037
038	81243	14947	84726	72591	91250	72145	43569	85013	12819	02383	038
039	75082	59280	58330	86168	40008	55288	67574	70079	21568	83130	039
040	01339	53335	77348	42907	68413	23632	06823	01132	65396	84563	040
041	02805	19078	85671	79317	08278	87149	78061	75300	79273	19950	041
042	52260	32477	36082	60922	60178	37754	22329	82588	19644	32605	042
043	94298	84308	48542	48022	77923	65827	62117	75816	02177	21504	043
044	04063	52653	62153	65990	52546	52666	62661	60477	92273	15574	044
045	60557	26087	61798	71916	07937	34938	34358	26830	10925	30028	045
046	52487	85247	38518	10041	34377	89337	25502	62838	02468	59466	046
047	60247	05726	41015	46428	27525	56463	75710	08503	81473	99271	047
048	87479	38192	05982	14800	79962	77007	28915	07529	89814	05564	048
049	59556	34367	30314	12900	75567	40169	28778	66663	80413	31159	049
050	20999	43175	47958	31740	05231	43299	82497	77328	13981	82061	050

	1	2	3	4	5	6	7	8	9	10	
051	72201	41294	56688	54267	69696	38928	66206	86275	56495	91473	051
052	51646	43481	07105	21706	27183	49286	71153	43293	43408	98949	052
053	20049	59233	01948	74735	84886	79630	21046	84127	48800	47896	053
054	13376	23060	27705	30523	49191	29449	20929	83007	64997	34667	054
055	72820	37114	15403	20453	59487	31630	12055	44870	95127	06361	055
056	82660	66526	59592	46438	42906	94220	25647	16358	39429	66873	056
057	66232	19463	46519	49495	00490	97011	00638	95504	22017	09711	057
058	45660	84124	86196	36733	85372	18118	52722	95632	43949	67989	058
059	83843	45857	45091	46024	62474	18757	90060	72398	55552	86798	059
060	26674	49635	45666	42739	00026	54091	00794	66239	76804	64374	060
061	56866	72236	08697	15059	55450	91735	53235	02273	64494	65759	061
062	72170	13581	64230	94335	04139	53915	45699	33165	64053	36333	062
063	38542	32773	27215	03450	25832	31468	84124	41717	52515	26407	063
064	96499	99270	37077	63685	22883	29302	44435	54460	67706	48573	064
065	64013	16885	58119	43152	65502	06484	73126	47608	10969	72049	065
066	03299	41478	19039	16434	11003	38622	26933	10435	11536	03824	066
067	06862	86524	89627	00791	28609	34555	61596	35367	08005	03371	067
068	40443	53595	53066	14362	48122	57775	82534	31723	95713	10679	068
069	57815	06496	08876	17550	21978	93104	85412	65727	15828	56495	069
070	92289	79363	54325	74606	28752	02572	76087	93884	67908	61765	070
071	42243	18779	17731	81639	13719	28966	38720	21185	89975	78479	071
072	29137	19152	56131	06478	81395	50385	99850	88972	60784	30443	072
073	70984	99223	70971	54536	12965	86838	92726	17259	69333	93505	073
074	15604	30256	20336	63350	73924	69261	69891	65975	96823	03620	074
075	37472	40894	64306	84969	48329	12175	78548	22857	62469	34813	075
076	59605	49429	46080	52721	22085	73764	36361	19185	61590	61261	076
077	75868	22254	48899	18491	08752	79381	50962	94732	60588	12600	077
078	94236	37642	91175	33539	09974	16366	18255	24952	74597	70431	078
079	79254	13250	22805	90364	83924	86580	50220	67924	89394	94991	079
080	80322	38740	73052	99094	45817	55755	38794	58927	12505	20960	080
081	24205	80349	08495	64180	19053	82901	77977	80786	29141	54205	081
082	62021	59292	12045	92351	33045	24430	81237	77352	98813	31889	082
083	41813	85501	98858	18096	33452	17507	44505	84588	08244	91264	083
084	33139	92458	89383	68157	26876	52086	71466	30088	40625	40887	084
085	80539	22138	93194	55278	40723	70350	57395	73836	16694	73799	085
086	68569	85866	40258	62403	09544	70935	99522	30405	95483	65022	086
087	11338	13818	58186	25776	18302	60468	60715	56740	52071	59561	087
088	85163	77680	88669	67548	36566	21664	44310	25260	94484	15458	088
089	43573	59076	65915	42561	63236	76076	13686	78072	39566	75503	089
090	87167	64197	30533	37308	27944	31128	20695	28097	82448	44275	090
091	63360	40091	91153	20036	68217	56249	61079	67304	61075	24853	091
092	89220	20649	63232	04651	93897	29505	05527	27652	08588	16026	092
093	54924	72707	84237	53755	98164	98661	90279	88739	96274	40071	093
094	26616	69868	52639	66750	90701	30715	40525	42275	91571	24772	094
095	99900	80123	24493	12438	23320	01350	10548	98347	99276	16789	095
096	98493	58870	56708	11212	41072	54287	72735	95649	66991	27123	096
097	47477	85534	86274	69753	43043	69242	94321	20909	43429	13601	097
098	87529	37934	70185	11735	10101	53137	08488	42215	84345	32343	098
099	18409	81970	20066	48638	72852	54930	07163	88619	56075	09736	099
100	14831	06586	13791	47309	72998	55657	57522	82007	57674	15703	100

	1	2	3	4	5	6	7	8	9	10	
101	07240	75212	78429	00618	69668	91210	22522	70024	85214	68992	101
102	26590	66591	47371	79670	66001	29693	93344	49725	31020	93530	102
103	97651	48672	59457	55097	85713	42405	20784	13221	03206	98993	103
104	20224	73798	90676	09967	16184	21871	76430	09296	86650	36842	104
105	72418	94098	63547	10195	98064	98096	17827	15232	67781	42414	105
106	07295	39405	93311	40944	44800	02897	66602	93894	13565	24782	106
107	58957	28744	47107	92790	24601	88598	39068	31210	63260	30978	107
108	60237	59247	70825	48103	53301	89335	47166	15386	53518	12307	108
109	67579	87183	28583	86544	65521	87922	81659	15142	01813	80038	109
110	80539	37623	85798	02580	38199	21453	68781	21501	69946	21351	110
111	41682	18953	34258	25173	52875	64839	11415	42235	54027	20124	111
112	70565	05231	03288	84448	16695	60321	22684	90580	32315	60997	112
113	53909	64871	54562	05027	52789	17292	18050	24424	54248	50166	113
114	12447	89558	72434	50596	68197	93275	90468	46127	24552	42265	114
115	13275	09467	63934	31891	66175	70613	15962	69399	24177	18392	115
116	37740	55614	72879	60917	67985	72684	18147	17516	71293	64315	116
117	73196	07641	58333	95529	01567	93087	98480	12760	39061	95912	117
118	38053	85351	21394	86178	74320	31578	22927	41388	56672	64846	118
119	90407	50719	69401	05134	56863	03227	45389	14471	71659	01597	119
120	65641	57244	91043	70827	04974	01837	64031	01416	27567	57726	120
121	42218	77378	31979	93614	78160	96480	36881	91947	67539	62415	121
122	69503	08871	96293	28518	16650	81337	17751	60444	09002	21094	122
123	63666	57422	78671	16138	04734	83378	30121	61584	09108	53557	123
124	52761	39774	18621	53134	97002	91979	62682	72029	76896	99900	124
125	88014	51530	66450	53950	08031	13066	39622	69529	14665	85055	125
126	60156	23091	33784	04975	89556	64945	70637	93681	31630	07592	126
127	65327	59339	86953	21184	20752	53776	75777	68888	43215	99963	127
128	83754	91957	19858	96846	08765	08547	21929	97387	60033	21352	128
129	01424	33234	36424	41401	18817	58695	29257	55481	04330	00439	129
130	83172	70152	89334	99489	24462	49999	69961	76475	43752	67592	130
131	92656	95733	74721	23629	71636	41044	79699	29068	12763	29562	131
132	15932	47873	69664	51555	86893	13423	50946	99842	66652	55924	132
133	23927	80033	46169	63637	69357	30673	36989	02847	62943	24637	133
134	90835	65481	71776	94827	27652	73888	34137	41467	20146	99080	134
135	15936	07031	06640	79899	12447	01310	17533	29745	76755	96126	135
136	75937	38163	17809	76671	10254	54745	17077	99910	77766	89416	136
137	63807	79646	31299	15290	61475	79718	28873	02654	26512	05597	137
138	91840	21890	38224	22321	71692	08929	36152	39832	17198	65103	138
139	57480	22066	82251	99664	71163	48268	08390	80071	38331	28790	139
140	81405	16278	27457	42043	00888	71262	29917	56739	52362	99326	140
141	91988	50456	46360	56483	26824	38045	15660	17994	88367	74828	141
142	23816	50670	64333	98238	84223	76482	42214	66837	37022	92566	142
143	56054	06285	38507	91559	83973	36482	18335	80724	00149	67022	143
144	94041	43796	19289	11220	14593	03383	54795	88413	92337	66809	144
145	76695	63320	27834	51387	84926	40090	65961	74089	45914	33491	145
146	15882	39007	65292	90502	82166	49323	46293	75517	82556	01030	146
147	03048	85241	56804	21259	56537	71967	22975	79426	83719	20350	147
148	75753	62032	18362	56662	94878	92137	69955	61268	50783	29753	148
149	42573	45655	15397	22513	45616	82713	61885	93170	17956	96806	149
150	43633	48300	59897	58554	16639	04950	24877	71723	21772	79051	150

	1	2	3	4	5	6	7	8	9	10	
151	43480	92280	28154	02900	22529	28510	04183	60759	36304	73895	151
152	12194	40230	17588	07593	62471	19845	94712	27406	31823	58443	152
153	21439	11824	30734	47613	23193	81366	74194	39207	95986	43388	153
154	20845	28517	66501	30212	61127	33712	98283	68917	41215	74833	154
155	64751	21384	54545	20699	97717	10202	46595	36513	23655	30244	155
156	27094	67239	95528	18581	28250	49321	88417	95669	74872	43988	156
157	34694	18239	69095	71908	38335	80393	49060	40678	23633	31119	157
158	57737	44362	19439	21203	66592	64802	92914	54168	81572	32662	158
159	82889	66827	93255	80141	66292	44643	13815	59627	04991	39540	159
160	08576	30416	28437	56809	29137	33128	08002	05832	07020	06744	160
161	35261	92162	18307	31914	23626	86095	09366	97881	51737	03669	161
162	15196	09876	70975	37617	29138	64422	15090	28131	64014	44261	162
163	68207	54875	49926	82937	36829	12361	23249	85348	92640	45471	163
164	82204	08460	27269	16268	22101	66875	66489	17250	60542	95383	164
165	29244	29145	99540	12759	85640	04695	03248	15209	73869	65420	165
166	80988	33891	12327	98904	03737	77086	04646	05694	64642	75827	166
167	61321	52265	15895	01124	47523	66313	12543	10104	23402	01062	167
168	14472	18698	40110	99995	69007	60168	12749	22455	99137	39379	168
169	79070	18512	73219	88390	91333	10563	94404	98730	33409	97501	169
170	83779	14573	47031	10558	01795	37822	92893	45749	67288	76381	170
171	01149	94643	06641	14567	55423	10957	27178	47082	15573	52799	171
172	06414	27054	88800	21004	82257	95804	29881	38526	45440	41808	172
173	81810	34212	26725	81655	50815	24927	55374	62099	73433	11642	173
174	15357	91786	44232	12599	28657	77480	43042	76221	05879	27958	174
175	79688	28910	15466	61878	92147	97375	31985	94768	70408	38208	175
176	46970	74041	05705	41250	52252	18772	59404	25092	84392	05570	176
177	31086	66346	51509	72443	59671	55418	33582	02805	54564	44624	177
178	30045	08506	51427	85922	07174	26677	15016	43235	89623	99158	178
179	46211	90723	03876	81275	86586	36915	28087	45978	92663	75896	179
180	05686	81389	28625	15902	06511	68733	48187	93629	96487	64654	180
181	25454	96934	47445	81382	76729	08238	65614	90516	77906	75417	181
182	76810	55665	38609	76297	69851	16939	51874	75354	54769	57350	182
183	89822	95857	49509	73093	41038	41579	78678	72042	72937	33560	183
184	16330	18407	17755	97418	05898	48168	94294	73042	45037	35833	184
185	62770	30212	97628	08356	56548	76128	14933	75910	12883	09662	185
186	15368	30667	57583	12359	39951	51863	72984	18629	76917	62365	186
187	06804	79498	99057	25197	46349	84507	54215	85986	20767	72154	187
188	34396	37904	63668	95631	35512	77182	04745	69043	55398	80699	188
189	10793	50005	69488	98464	50558	00868	58666	47049	67205	86498	189
190	13773	41480	40390	93421	84049	88753	31969	08777	40582	15060	190
191	84955	39456	58582	30198	89726	85610	96026	49798	51512	36518	191
192	39288	65123	61678	48804	86229	60887	55153	74117	87115	14440	192
193	57021	67906	90450	33491	05562	10318	30062	55912	84330	54436	193
194	82430	45734	40166	51637	46118	76745	86389	80162	25199	98885	194
195	56922	64973	21579	41910	39697	55701	63379	80391	68987	11256	195
196	52346	84772	44011	10682	07809	57379	44263	88684	75688	80789	196
197	97174	53700	70906	37204	94738	12572	43938	13446	25197	72526	197
198	66985	78370	04055	24812	67527	05695	45035	81358	62751	82865	198
199	56619	20435	47570	43166	38184	21804	50939	63887	25300	01590	199
200	84351	43554	14397	37947	43603	87859	94393	42928	28817	95920	200

CONTENTS OF A NON-INTEGRATED SAMPLING PLAN

I. INTRODUCTION

This section will serve as an overview for quick reference. At a minimum, it should include a paragraph similar to the following example:

> This sampling plan serves as the foundation for the STATE of ________________'s Food Stamp Quality Control Sampling Procedures for Fiscal Year ______. Based on the projected average monthly caseload of _________ in the active universe we elect to review the minimum of _____ cases. The active frame will be constructed from the _____________ and the interval used will be _____. Based on the negative average monthly caseload figures of _______ the minimum of _____ cases will be selected for review using an interval of ___.
>
> Both samples will be selected using systematic random sampling without integration or planned stratification.

If the design contains stratification or integration, describe each stratum as above.

II. SAMPLING FRAMES

This section shall include a complete description of both the active and negative sampling frames. Address each frame separately. At a minimum, discuss the following items:

A <u>Sources</u>: Sources of the frames (such as lists of redeemed ATP's, mail issuance records, daily tally sheets, list of certified eligibles). If more than one source is used to compile the frame, describe the method of selection from each part of the frame.

B <u>Availability</u>: The times at which the frames, including any segments and extensions of them, are available for sample selection. Explain the monthly selection cycle and the time schedules involved.

C <u>Accuracy and Completeness</u>: The accuracy and completeness of sample frames compared to the target populations. Explain the checks-and-balance methodology used to verify and reconcile the frame figures.

D <u>Components</u>: Whether or not the frame are constructed by combining more than one list. Describe the process and time frames involved in constructing the list.

E <u>Location</u>: Where the frames are constructed and maintained, such as entirely in a central State office, entirely in local offices, or in the State office based on information supplied by local offices.

F <u>Form</u>: The form of the frames, such as computer files, microfilm, or hard copy. If different parts of the frame are in different forms, the form of each part must be specified.

G <u>Frequency of Updates</u>: Describe the time frames associated with and the frequency of updating the frames or their sources.

H <u>Deletion of Cases Not Subject to Review</u>: Methods of identifying, locating, and deleting cases not subject to review.

I <u>Structure</u>: Describe the structure of the frames. For computerized frames in physical-sequential files, describe the order of cases within each frame, both prior to the selection process and any sorting involved afterwards. For computerized frames in random-access files, describe the access keys used in frame construction and sample selection.

III. SAMPLE SELECTION

Describe the method of sample selection by discussing the following items:

A <u>Caseload Estimation</u>: Show the actual computations you used in estimating reviewable caseloads. (Include computations for the out-of-scope rate.) In particular, include outputs from all associated computer programs and specify formulas for and include outputs from all spreadsheet calculations.

B <u>Sample Sizes</u>: Specify and explain the basis for your minimum sample sizes.

If you opt for a reduced active sample, include a waiver statement in which (regardless of whether or not you increase your sample size to reflect an increase in caseload) you agree not to contest the final error rates and sanctions on the basis of the precision resulting from the reduced sample size. (See Exhibit G for an example of a waiver statement.)

If the sample design is not a simple-random sample or systematic sample, demonstrate that the design would yield estimates of the payment error rate or negative error rate with same or better precision as would be obtained by simple random samples of the sizes specified in 3121.1.

Show the actual computation of the target sample size, with the projection of the overselection factor based on historical sample data.

C <u>Computation of Sampling Intervals and Random Starts</u>: Include the methodology and computations for both; plus, include computations for the out-of-scope rate. (Applies only to systematic sampling.)

D <u>Stratification or Clustering</u>: Explain in detail the criteria and operational environment involving stratification or clustering. This includes the characteristics of each strata or cluster.

E <u>Sample Selection Procedure</u>: Explain your sample selection procedures and your numbering system for sampled cases. Explain the relationship, if any, between the food stamp QC sample design and other federally mandated quality control samples (e.g.,AFDC or Medicaid).

F <u>Corrections for Over- or Undersampling</u>: State your procedure (or the set of procedures you will choose from to compensate for excessive oversampling or undersampling (e.g., reserve pools, sample adjustments during the year and at the end of the year, or stratification by time satisfying the equal-or-better precision requirement).

G <u>Monitoring of Sample Selection and Assignments</u>: If your sample is manually selected, describe your checks and controls to monitor the selection and assignment of the sample cases.

EXAMPLE OF FORM FNS-248 AND INSTRUCTIONS FOR COMPLETION

FORM APPROVED OMB NO. 0584-0034

U.S. DEPARTMENT OF AGRICULTURE - FOOD AND NUTRITION SERVICE **STATUS OF SAMPLE SELECTION AND COMPLETION** FNS HANDBOOK 311	STATE	FISCAL YEAR MONTH

Public reporting burden for this collection of information is estimated to average 30 minutes per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Agriculture, Clearance Officer, OIRM, Room 404-W, Washington, D.C. 20250; and to the Office of Information and Regulatory Affairs, Office of Management and Budget, Washington, D.C. 20503.

		STRATUM ____		STRATUM ____		STRATUM ____		STRATUM ____	
		1a	1b	2a	2b	3a	3b	4a	4b
I. ACTIVE CASE INFORMATION	LINE NO.	REPORTING MONTH	YEAR TO DATE	REPORTING MONTH	YEAR TO DATE	REPORTING MONTH	YEAR TO DATE	REPORTING MONTH	YEAR TO DATE
A. TOTAL FRAME SIZE	1								
B. TARGET SAMPLE SIZE	2								
C. SAMPLING INTERVAL	3								
D. SAMPLE CASES SELECTED	4								
II. NEGATIVE CASE INFORMATION		STRATUM ____ 1a	1b	STRATUM ____ 2a	2b	STRATUM ____ 3a	3b	STRATUM ____ 4a	4b
A. TOTAL FRAME SIZE	5								
B. TARGET SAMPLE SIZE	6								
C. SAMPLING INTERVAL	7								
D. SAMPLE CASES SELECTED	8								
1. Pending	9								
2. Disposed	10								
a. Not Subject to Review	11								
b. Not Completed	12								
c. Completed	13								
1. No Error	14								
2. Error	15								

SPECIMEN

NOTES:

FORM FNS-248 (11-89) *Previous edition obsolete*

K100 COLUMN INSTRUCTIONS.

If your sample is disproportionately stratified, enter individual-stratum values in separate columns. If the number of strata exceeds four, use additional forms. Enter the stratum identifiers in the column headings.

The following instructions for columns 1a and 1b also apply to columns 2 through 4.

A Reporting Month (Column 1a). Information for each reporting month shall be entered in column 1a. Entries for the report month reflect the activity for the sample month and not necessarily activities during the calendar month. For example, if you selected 100 cases for the sample month of January but actually completed 110 during the calendar month of January, then the dispositions of the 100 January sample-month cases would be reported on the January Form FNS-248.

B Year To Date (Column 1b). Over the course of the year accumulate monthly entries from column 1a into the adjacent entries in column 1b. Also, update column 1b for for any cases that (1) were in pending status in prior months but now are no longer pending or (2) have had a change in disposition or finding. Also, update column 1b to account for cases coming from a supplemental sample or selected from a reserve pool that you add to the prior months to correct for undersampling. For example, it is time to complete the Form FNS-248 for the report month of May. However, the April report reflected 10 active cases as pending and two active cases as not completed, and some time after the submission of the April report, these 12 cases were completed. Then their respective disposition changes would be reflected in column 1b of the May report.

Complete column 1b for all lines, except lines 3 and 7 (active- and negative-case sampling intervals). Column 1b only includes the dispositions for cases selected for months prior to and including the report month. For example, it is now time to complete the Form FNS-248 for the report month of February. Column 1b would include the disposition totals for months October through February only. You may have some case dispositions for March or beyond at the time of the report preparation, but they are not to be included in the report for the February reporting month.

K200 LINE-BY-LINE INSTRUCTIONS FOR ACTIVE CASES. Lines 1 through 4 apply to active cases.

A Total Frame Size. (Line 1). Enter the actual number of active cases that you subjected to sampling. Do not estimate this number but show the total frame size for the sample month/reporting period. If within the sample month your sample is disproportionately stratified, enter the total frame size for each stratum.

B Target Sample Size. (Line 2). Enter the number of cases that are projected as being selected for this sample month/reporting period. This figure will include the overselection factor to compensate for cases that are selected but are not subject to review. If within the sample month your sample is disproportionately stratified, enter the target sample size for each stratum.

For example, assume your minimum sample size is 1200. Historically, you have noticed that your proportion of cases that are not subject to review is 4.00%. Therefore, your projected number of cases to select for the year would be 1250. (1200 divided by 0.96 equals 1250). Your monthly target sample size would be at least 105 (104 would not yield enough cases). In column 1b, you would enter 105 on the October report, 210 on the November report, 315 on the December report, etc.

C Sampling Interval. (Line 3). In column 1a, enter the sampling interval(s) you used to select those active cases that were to be reviewed for the report month. (If you performed reserve-pool subsampling, enter the effective interval (see 3122(B).) If within the sample month your sample is disproportionately stratified, enter the intervals for each stratum. If the interval has changed during the course of the year, place an asterisk in column 1b and describe in a footnote at the bottom of the form the history of the interval changes.

D Sample Cases Selected. (Line 4). Enter the total number of active cases selected from the sample frame and/or from a reserve pool. If within the sample month your sample is disproportionately stratified, enter the number of cases selected from each stratum. Exclude selected cases you have put into a reserve pool but have not yet withdrawn them from the pool. If the entries in either column 1a or 1b contain cases selected from a reserve pool, provide in footnotes to these entries a breakdown of the number of cases selected from the sampling frame versus the number selected from a reserve pool. In specifying the latter, if you have a stratified sample, breakdown the reserve-pool selection by stratum.

K300 LINE-BY-LINE INSTRUCTIONS FOR NEGATIVE CASES.

Lines 5 through 15 apply to negative cases.

A Total Frame Size. (Line 5). Enter the actual number of negative cases that you subjected to sampling. Do not estimate this number but show the total frame size for the sample month/reporting period. If within the sample month your sample is disproportionately stratified, enter the total frame size for each stratum.

B Target Sample Size. (Line 6). In column 1a, enter the projected sample size for the report month. This includes the overselection factor for cases that are selected but are not subject to review. If within the sample month your sample is disproportionately stratified, enter the target sample size for each stratum. In column 1b, report the accumulation of the monthly target sample sizes from the beginning of the year through the report month.

C Sampling Interval. (Line 7). In column 1a, enter the sampling interval you used to select the negative cases that were to be reviewed for the report month. (If you performed reserve-pool subsampling, enter the effective interval (see 3122 B.) If within the sample month your sample is disproportionately stratified, enter the intervals for each stratum. If the interval has changed during the course of the year, place an asterisk in column 1b and describe in a footnote at the bottom of the form the history of the interval changes.

D Sample Cases Selected: Total. (Line 8). Enter the total number of negative cases selected from the sample frame and/or from a reserve pool. If within the sample month your sample is disproportionately stratified, enter the number of cases selected from each stratum. Exclude selected cases you have put into a reserve pool but have not yet withdrawn them from the pool. If the entries in either column 1a or 1b contain cases selected from a reserve pool, provide in footnotes to these entries a breakdown of the number of cases selected from the sampling frame versus the number selected from a reserve pool. In specifying the latter, if you have a stratified sample, breakdown the reserve-pool selection by stratum.

1 Sample Cases Selected: Pending. (Line 9). Enter the number of negative sample cases with pending reviews as of the end of the report month/reporting period. These are cases that have not been reviewed, and no disposition has been assigned. They are not to be confused with line 12 (Not Completed).

2 Sample Cases Selected: Disposed. (Line 10). Enter the number of disposed negative cases. This line plus line 9 (pending) must equal line 8 (total).

a Disposed: Not Subject to Review. (Line 11). Enter the number of cases coded disposition 2.

b Disposed: Not Completed. (Line 12). Enter the number of cases in the negative sample for which a review was not completed and therefore coded 3 or 4.

c <u>Disposed: Completed. (Line 13)</u>. Enter the number of cases that have been completed. This does not include those in pending status or those that have been coded Not Subject to Review or Not Completed.

i <u>Completed: No Error. (Line 14)</u>. Enter the number of cases from line 13 (completed) that do not contain an error. These are the cases in which the review was completed, and the household was found to be correctly denied or terminated.

ii <u>Completed: Error. (Line 15)</u>. Enter the number of cases from line 13 (completed) that contained an error finding. These are cases in which the review was completed, and the denial or termination action was found to be invalid. Line 14 plus line 15 must equal line 13.

EXAMPLE OF FORM FNS-247 AND INSTRUCTIONS FOR COMPLETION

FORM APPROVED OMB NO. 0584-0034

U.S. DEPARTMENT OF AGRICULTURE – FOOD AND NUTRITION SERVICE STATISTICAL SUMMARY OF SAMPLE DISPOSITION FNS HANDBOOK 311	STATE	FISCAL YEAR

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Agriculture, Clearance Officer, OIRM, Room 404-W, Washington, D.C. 20250; and to the Office of Information and Regulatory Affairs, Office of Management and Budget, Washington, D.C. 20503.

I. ACTIVE CASE INFORMATION	LINE NO.	STRATUM ___ INTERVAL ___	STRATUM ___ INTERVAL ___	STRATUM ___ INTERVAL ___	STRATUM ___ INTERVAL ___
A. ACTUAL AVERAGE MONTHLY CASELOAD SUBJECT TO REVIEW	1				
B. END OF YEAR MINIMUM SAMPLE SIZE	2				
II. NEGATIVE CASE INFORMATION	**LINE NO.**	**STRATUM ___ INTERVAL ___**	**STRATUM ___ INTERVAL ___**	**STRATUM ___ INTERVAL ___**	**STRATUM ___ INTERVAL ___**
A. ACTUAL AVERAGE MONTHLY CASELOAD SUBJECT TO REVIEW	3				
B. END OF YEAR MINIMUM SAMPLE SIZE	4				
C. DISPOSITION OF SAMPLE CASES					
1. Number of cases selected	5				
2. Number of cases not subject to review	6				
3. Number of cases not completed	7				
a. Unable to locate case record or participant	8				
b. Review not processed	9				
4. Number of completed case reviews	10				
D. CASES IN ERROR	11				
1. Percent	12				

REMARKS:

SPECIMEN

FORM FNS-247 (2-90) *Previous edition obsolete*

L100 LINE-BY-LINE INSTRUCTIONS FOR ACTIVE CASES. Lines 1 and 2 apply to active cases.

A Actual Average Monthly Caseload Subject To Review. (Line 1). Enter for active cases the actual average-monthly caseload subject to review. (See 3320 for calculation formulas.)

B End of Year Minimum Sample Size. (Line 2). Enter for active cases your end-of-year minimum sample size. If your sample is unstratified or proportionally stratified, see 3121.1(B) for calculation formulas. If your sample is disproportionately stratified, do not complete this entry at the stratum level; instead, calculate an entire-sample entry using the formulas in E330.

L200 LINE-BY-LINE INSTRUCTIONS FOR NEGATIVE CASES. Lines 3 through 12 apply to negative cases.

A Actual Average Monthly Caseload Subject to Review. (Line 3). Enter for negative cases the actual average-monthly caseload subject to review. (See the calculation formulas in 3320 and the work sheet in Figure L-1.)

B End of Year Minimum Sample Size. (Line 4). Enter for negative cases your end-of-year minimum sample size. If your sample is unstratified or proportionally stratified, see 3121.1(B) for calculation formulas. If your sample is disproportionately stratified, do not complete this entry at the stratum level; instead, calculate an entire-sample entry using the formulas in E330.

C Disposition of Sample Cases.

1 Number of Cases Selected. (Line 5). Enter the number of cases that you selected for negative case reviews. Include cases that were completed, not completed, or not subject to review. This entry equals the sum of lines 6, 7, and 10.

2 Number of Cases Not Subject to Review. (Line 6). Enter the number of cases coded with a disposition of 2. This count includes cases deselected by a correction for oversampling.

3 Number of Cases Not Completed. (Line 7). Enter the number of cases selected for negative case reviews but were later coded as not completed. This is the sum of all cases with a disposition code or 3 or 4. This line equals the sum of lines 8 and 9.

a Unable to locate Case Record or Participant. (Line 8). Enter the number of cases with a disposition of 3.

Figure L-1

Work sheet for negative cases to calculate (1) average-monthly caseload subject to review and (2) weighted negative error rate for disproportionately stratified samples. Instructions:

A Complete lines 5, 6, 10 and 11 of the negative-case Form FNS-247. If you have a disproportionately stratified sample, do this for each stratum. In the below table, put line 5 entries in column C, line 6 in column D, and the ratio of line 11 to line 10 in column G.

B Fill in column B with the value of I determined by equation (3-11) or by your sampling interval (see 3320).

C Calculate the entry for column E by using the formula shown at the top of column E.

D Add up the entries in column E and record total at bottom of column E.

E Calculate weights in column F using the formula shown at the top of column F.

F Calculate entries for column H using the formula shown at the top of column H. Add up the entries in column H to give the weighted negative error rate.

G Obtain the average-monthly subject-to-review caseload by dividing column E by the number of months that the stratum is in effect.

A	B	C	D	E	F	G	H
stratum	I	# cases selected LINE 5	# cases NSR LINE 6	Compute: B*(C-D)	(Weights) Compute: E/ETOTAL	Unweighted % in error LN11/LN10	Compute: F*G
						%	
						%	
						%	
						%	
						%	
						%	
			ETOTAL:		Weighted Line 12:		%

b Review not Processed. (Line 9). Enter the number of cases with a disposition of 4.

4 Number of Completed Cases Reviews. (Line 10). Enter the number of cases selected in your negative sample for which a review was completed. This is a count of the number of cases with a disposition of 1.

D Cases in Error. (Line 11). Enter the total number of negative cases for which a review was completed and the household was found to be incorrectly denied or terminated. The percent (Line 12) is calculated by dividing line 11 by line 10. Express the entry for line 12 as a percentage and retain two decimal places, e.g. 3.45. If your sample is disproportionately stratified, also calculate the weighted entry for line 12 that estimates the negative error rate for the entire population. Figure L-1 contains a work sheet for doing this.

WORK SHEET FOR DETERMINING THE EFFECTS OF A CASELOAD CHANGE

A. On the lines below enter the values of the indicated quantities:

symbol	description	value
Ne	Estimated average-monthly caseload subject to review	______
ne	Minimum sample size based on Ne	______
I	Sampling interval	______
Nu	Updated average-monthly caseload subject to review (see 3411.1)	______

B. On lines 1.1 and 2.1 below enter the values of 1.2*Ne and Ne, respectively. Place an "X" on line 1, 2, or 3 to indicate the relationship of Nu with the entered values:

	____________	(1) :	"X" if Ne > (1.1)
1.2 * Ne =	____________	(1.1)	
	____________	(2) :	"X" if (1.1) > Ne > (2.1)
Ne =	____________	(2.1)	
	____________	(3) :	"X" if (2.1) > Ne

For example, if the updated caseload, Nu, is less than Ne, then place an "X" on line 3.

C. If you placed an "X" on line 2, your minimum sample size remains unchanged; set the updated minimum sample size, nu, equal to ne and go to step D. Otherwise, calculate nu using the appropriate formula below:

ACTIVES	Nu<10,000	nu = 300	____
(non-reduced)	Nu>60,000	nu = 2400	____
	otherwise	nu = 300 + .042(Nu − 10,000) =	____
(reduced)	Nu>60,000	nu = 1200	____
	otherwise	nu = 300 + .018(Nu − 10,000) =	____
NEGATIVES	Nu<500	nu = 150	____
	Nu>5000	nu = 800	____
	otherwise	nu = 150 + .144(Nu − 500) =	____

D. If you have an unstratified (or proportionally stratified) systematic sample, estimate how many reviewable cases, na, would be selected if you continued using the current sampling interval:

$$na = (12*Nu) / I = \_\_\_\_\_\_\_\_.$$

For all other types of samples, estimate how many reviewable cases, na, would be selected if you continued using the current sampling parameters:

$$na = S1 + S2 = \_\_\_\_\_\_\_\_ + \_\_\_\_\_\_\_\_ = \_\_\_\_\_\_\_\_$$

where

S1 = the number of cases selected (and not deselected) that were subject to review in the preceding m months (This is determined by adjusting the number of selected (but not deselected) cases downward by the anticipated number of Not-Subject-to-Review cases.), and

S2 = the predicted number of reviewable cases to be obtained in the remaining 12-m months, assuming no sampling corrections.

If you believe the first m months are representative of the entire year, use

$$na = (12)(S1)/m = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ .$$

E. Determine the difference between the revised minimum sample size, nu, computed in step B or C and the projected sample size, n_a, computed in step D:

$$DIFFERENCE = na - nu = \_\_\_\_\_\_\_\_.$$

F. If DIFFERENCE is positive, then you are oversampling. A negative value for DIFFERENCE indicates undersampling. Depending on the magnitude of the SURPLUS or DEFICIT and how many more months remain to be sampled, make a decision on whether to correct for under- or oversampling. It may be preferable to postpone making a correction until later in the year when a trend may be more definite.

BIBLIOGRAPHY

100 FNS REGULATIONS. The following FNS regulations state the Food Stamp Program requirements for State sampling procedures and associated activities:

A Regulation 7 CFR 272.1(f)--retention of records,

B Regulation 7 CFR 272.2(e)(4)--submittal requirements for sampling plans,

C Regulations in 7 CFR Part 275--Performance Reporting System, and

D Regulation 7 CFR 276.4--suspension/disallowance of administrative funds

200 SAMPLING TEXTBOOKS.

Cochran, William G. (1977) Sampling Techniques. Third edition, John Wiley and Sons, New York.

Hansen, M.H., Hurwitz, W.N., and Madow, W. G. (1953) Sampling Survey Methods and Theory (2 Vols.). John Wiley and Sons, New York.

Jessen, Raymond J. (1978) Statistical Survey Techniques. John Wiley and Sons, New York.

Kalton, Graham (1983) Introduction to Survey Sampling. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-035, Sage Publications, Beverly Hills, California.

Kish, Leslie (1965) Survey Sampling. John Wiley and Sons, New York.

Kish, Leslie (1987) Statistical Design for Research. John Wiley and Sons, New York.

Konijn, H.S. (1973) Statistical Theory of Sample Survey Design and Analysis. American Elsevier Publishing, New York.

Mendenhall, W., Ott, L., and Scheaffer, R. (1971) Elementary Survey Sampling. Wadsworth.

Raj, Des (1968) Sampling Theory. McGraw-Hill, New York.

Sudman, Seymour (1976) Applied Sampling. Academic Press.

Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S., and Asok, C. (1984) Sampling Theory of Surveys with Applications. Third revised edition, Iowa State University Press, Ames, Iowa.

Yates, Frank (1981) Sampling Methods for Censuses and Surveys. Fourth edition, Oxford University Press, New York.

300 PREDICTION OF CASELOADS.

Barnow, Bart, and Garasky, Steven (1988) "The Use of Regression Modeling to Evaluate the REACH Program in New Jersey". (National Association for Welfare Research and Statistics; Proceedings of the 28th National Workshop on Welfare Research and Statistics; Baltimore, Maryland; July 25-28, 1988; pp. 127- 130).

Plotnick, Robert, and Lidman, Russell (1986) "Forecasting AFDC Caseloads in Washington State". (National Association for Welfare Research and Statistics; Proceedings of the 26th National Workshop on Welfare Research and Statistics; Richmond, Virginia; July 27-30, 1986; pp. 44-59).

Williams, Deborah (1987) "Problems in Predicting Caseloads: Seasonal Data". (National Association for Welfare Research & Statistics; Proceedings of the 27th Annual Workshop on Welfare Research and Statistics; Louisville, Kentucky; August 3-6, 1987; pp. 70-77).

400 AUTOMATED METHODS FOR SAMPLE SELECTION AND PSEUDO-RANDOM NUMBERS.

Kahener, David; Cleve Moler; and Stephen Nash (1989). Numerical Methods and Software. Prentice Hall, New York.

Naylor, Thomas, et. al. (1968) Computer Simulation Techniques. John Wiley and Sons, New York.