Chapter Three: Deriving the Transition Matrix

The study's basic analytic construct is a five-by-five matrix of transition probabilities that indicate month-to-month patterns of food stamp participation and error. This chapter describes how the transition matrix is derived from available national data, through a five-by-five crosstabulation of households according to their food stamp status in one month and in the following month.

Exhibit 6 shows the generic notation used to identify the cell entries of the basic crosstabulation. Rows of the matrix correspond to current-month status; columns correspond to next-month status. Illustratively, as explained later, we regard the current month as March and the next month as April. The household count in row i and column j (the cell entry Q_{ij}) indicates the number of households in group i at the start of a month who then belong to group j at the start of the following month. In each month, every household is considered to belong in one (and only one) of the five specified groups.

Representing Food Stamp Administrative Procedures

This formulation of the model is chosen as a way of describing food stamp participation rates and error rates as outcomes of month-to-month household changes, with an explicit focus on initial certification, interim action, and recertification as distinct stages in the food stamp administrative process.

Initial Certification

The initial certification process is represented by the first row of the matrix. This row corresponds to households that at the start of the current month are not receiving food stamps but who may transition into the program in the subsequent month. For nonparticipating households at the start of the current month, their distribution at the start of the next month depends on whether: (a) they remain nonparticipants (Q_{11}), (b) they become correct cases (Q_{12}), or (c) they become error cases (Q_{13}). As noted above, we make the simplifying assumption that newly-certified cases are not assigned a one-month certification period; this implies that Q_{14} and Q_{15} are both zero.

Interim Action

The interim action process is represented by the second and third rows of the matrix. These rows correspond to ongoing food stamp cases—households that at the start of the current month are active cases not subject to recertification (i.e., not in the final month of their certification period), and are either correct (second row) or in error (third row). Consider the

Exhibit 6: Basic Transition Matrix

Current-month (“March”) status	Next-month (“April”) status: Non-participating	Ongoing correct	Ongoing error	Expiring correct	Expiring error	Total (row)
	Number of households					
Nonparticipating	Q_{11}	Q_{12}	Q_{13}	$Q_{14}=0$	$Q_{15}=0$	R_1
Ongoing correct	Q_{21}	Q_{22}	Q_{23}	Q_{24}	Q_{25}	R_2
Ongoing error	Q_{31}	Q_{32}	Q_{33}	Q_{34}	Q_{35}	R_3
Expiring correct	Q_{41}	Q_{42}	Q_{43}	$Q_{44}=0$	$Q_{45}=0$	R_4
Expiring error	Q_{51}	Q_{52}	Q_{53}	$Q_{54}=0$	$Q_{55}=0$	R_5
Total (column)	C_1	C_2	C_3	C_4	C_5	Q
Data source (see Notes below)	(a)	(b)	(b)	(b)	(b)	(c)

Notes:

a. First-column entries and total: calculated as row residuals.

b. Second-, third-, fourth-, and fifth-column entries and totals: estimated from current-year QC data.

c. Row totals: The values for R_1 through R_5 are derived on the basis of linear interpolation between the current-year and prior-year distribution of households, as explained in the text. The value for Q (total households in the population) is obtained from Census data for April of each year.

second row, the ongoing correct cases. In the next month such a household can either: (a) become a nonparticipating household (Q_{21}); (b) remain a correct case, not in its final certification month (Q_{22}); (c) become an error case, not in its final certification month (Q_{23}); remain a correct case, in its final certification month (Q_{24}); or become an error case, in its final certification month (Q_{25}). Similar month-to-month transitions can occur for ongoing error cases—i.e., error cases in the midst of a certification period. The corresponding third-row cell entries are Q_{31}, Q_{32}, Q_{33}, Q_{34}, and Q_{35}.

Recertification

The recertification process is represented by the fourth and fifth rows of the matrix. These rows correspond to expiring cases: cases that in the current month are in the final month of their certification period, either correct (fourth row) or in error (fifth row). Consider the fourth row—the expiring correct cases. The recertification about to be conducted on such a case will result in the household's either: (a) becoming a nonparticipating household (Q_{41}); (b) remaining correct (Q_{42}); or (c) becoming in error (Q_{43}). Similar month-to-month transitions can occur for expiring error cases—those incorrectly paid in the final month of their certification period (corresponding to Q_{51}, Q_{52}, and Q_{53}). We assume that cases subject to recertification do not become overdue for recertification and that newly-recertified cases are not assigned a one-month certification period (implying that Q_{44}, Q_{45}, Q_{54}, and Q_{55} are all zero).

Using National Food Stamp Quality Control Data

The primary data source for this analysis is the national Food Stamp Program Quality Control (FSPQC) system. Each year a nationally representative sample of between 45,000 and 50,000 active food stamp cases is selected for QC review, to assess the accuracy of eligibility and benefit determinations made by the administering state and local program agencies. Annual samples range from 300 to 2,400 by state, depending on the size of each state's average monthly active caseload. Approximately one-twelfth of a state's annual sample is drawn each month for review by state QC reviewers. The QC review consists of an inspection of the case record, an interview with the household, and additional field investigation that may include contacts with collateral sources such as employers, landlords, and banks. The information recorded by the state QC reviewers is assembled into a national QC database by USDA's Food and Nutrition Service. National and state error rates are calculated using this database.[15]

[15] A subsample of each state's QC sample is then subject to validation by federal QC staff, and the official error statistics incorporate the findings of both the state reviews and federal re-reviews. The federal re-review findings are not used in this study, as the federal subsample is about one-third the size of the full QC sample.

The decision to represent household transitions on a month-by-month basis reflects the fact that the active case QC data describe the status of cases in monthly terms. Specifically, the national QC database provides information about the presence of error for sample cases in their review month, along with information enabling one to deduce their participation and error status in the prior month. ***In the model developed here, we use the term "next month" to refer to the QC review month.*** The term "current month" refers to the month prior to the review month.

The model makes no assumption about whether the current-month distribution of households across the five groups (as indicated by the row totals, R_i) is the same as the next-month distribution (as indicated by the column totals, C_i). The expected situation, for the nation or for any particular state, is that the distribution of households is in short-term flux from one month to the next. The estimates derived in this analysis reflect such short-term month-to-month fluctuations. In general, the nature of a Markov model is such that the monthly transitions will (if uninterrupted) lead to a stable long-term distribution of households. The modeling approach here does ***not*** assume, however, that the nation or any individual state has already reached such an equilibrium.

As described below, the task of empirically deriving the model is a matter of using the national QC sample data on active food stamp cases (along with Census data on the total household population) to arrive at cell counts for the basic transition matrix. These cell counts, appropriately weighted to account for the QC sampling procedures, are then used to compute the transition probabilities. One can view the estimation as an accounting process whereby households are placed within the five-by-five classification according to their status in the review month and the preceding month. As described below, a number of assumptions are required to overcome the limitations of the QC data.

In principle, one would want to derive the model separately for each calendar month. This would mean, for instance, that one would calculate twelve separate monthly models over any given annual interval. The monthly sample sizes in the national QC data—approximately 4,000 cases reviewed per month—are not sufficient, however, to support the monthly derivation of the model, even at the national level. This is because some of the monthly transitions in question are based on small segments of the active caseload—for example, cases with earnings that are in error at the final month of their certification period. Moreover, some of the transitions in question are relatively rare events, occurring with a probability of less than 5 percent.

For these reasons, we have chosen to estimate the national model using the full QC sample for a federal fiscal year (October through September). ***We assume that the month-to-month pattern of change is the same in each of the twelve months of the fiscal year.*** The estimates for the year in question are thus not specific to any particular calendar month; there is no assumed intra-year variation in monthly error patterns.

Under this approach, ***we have found it helpful (although not necessary) to consider the data for each fiscal year as having been collected in the mid-year review month of April***. We thus interpret the month-to-month transitions as March-to-April transitions. In Exhibit 6, we show the current-month status as referring to "March" and the next-month status as referring to "April."

The active case QC data and Census data allow us to estimate directly some, but not all, cell counts in the matrix. Most notably, the number of cases that exit from the caseload each month is not observed in the active case QC data and must be estimated indirectly as a row or column residual. The resulting closure rates are subject to variation, as they reflect the sampling error of all elements entering the calculation. Exhibit 6 indicates which cell entries are estimated directly from QC data or indirectly as row or column residuals. ***The first-column entries of the matrix correspond to households that are not participating in the review month ("April"). They are not observed in the active case QC data. These cell counts thus cannot be directly computed and must be derived.*** If one knows the row totals of the matrix—that is, the preceding ("March") monthly count of households in each of the five subgroups—one can compute each first-column entry as a row residual (i.e., by subtracting all other row entries from the row total).

We have calculated the basic matrix of cell counts for each year 1998 through 2001 using national data from the food stamp QC reviews of active cases. We have excluded Guam and the Virgin Islands from the analysis, as the necessary Census data on annual household counts do not include the territories.[16]

Exhibit 7 shows the total number of food stamp QC cases included in the analysis, for 1997 through 2001.

[16] The annual QC data files, including details on error cases not included in public-use data sets, were made available to us by Mathematica Policy Research (MPR), at the request of the Food and Nutrition Service (FNS). We wish to acknowledge the assistance provided by Jenny Genser of FNS and Karen Cunnyngham of MPR in making the detailed data available for this analysis.

Exhibit 7: Size of the Analysis Sample, 1997-2001

	1997	1998	1999	2000	2001
	Number of food stamp quality control cases				
Households with earnings	11,968	12,460	12,925	12,683	12,459
Households without earnings	35,977	33,896	33,702	33,383	33,686
Total households	47,945	46,356	46,627	46,066	46,145

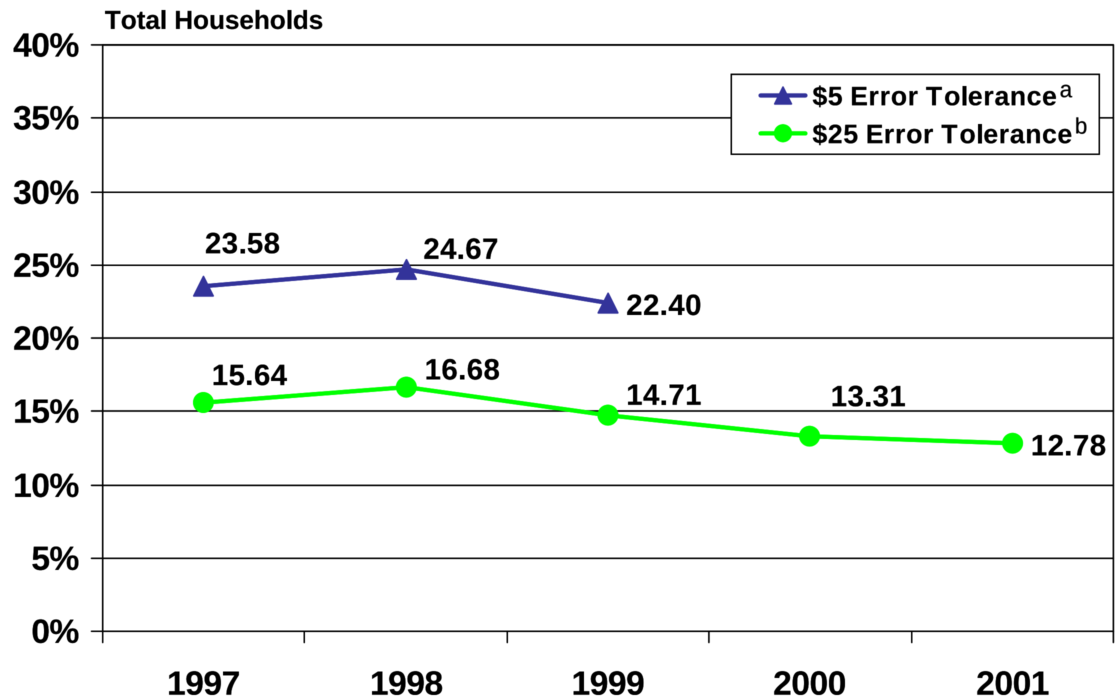
Note: Analysis sample excludes Guam and Virgin Islands.

Several adjustments to the data were necessary, as follows:

- To establish a consistent error standard for the historical period, we retroactively applied to the 1997, 1998, and 1999 data the $25 error tolerance for active cases that were eligible for food stamps. As noted earlier, this QC policy provision was first effective in 2000. Thus, for each of the years analyzed here, an eligible active case was considered correct if the monthly benefit differed from the correct amount by $24 or less. (There is no error tolerance for active cases that are ineligible.) Prior to 2000, the error tolerance for active cases was $5. Consistency across years required use of the $25 tolerance rather than the $5 tolerance, as it was not possible to apply the $5 tolerance to cases in 2000 and 2001. (We are unable to identify the eligible cases with error amounts of $5 to $24, as these cases were found correctly paid and no error amount was recorded.) As shown in Exhibit 8, approximately one-third of the cases classified as errors in 1997-1999 were eligible cases with overpayment or underpayment of $5 to $24. Such cases were reclassified as correct in this analysis, to establish a consistent error definition.

- We considered cases with a one-month certification period as having a two-month certification period. This was to avoid the calculation of some transition probabilities on the basis of very small samples. Note that less than 0.5 percent of all sample cases were assigned a one-month certification period. (Separately, two cases with an indicated certification length of 0 were excluded from the analysis.)

Exhibit 8: National Case Error Rates with $5 and $25 Error Tolerance, 1997-2001



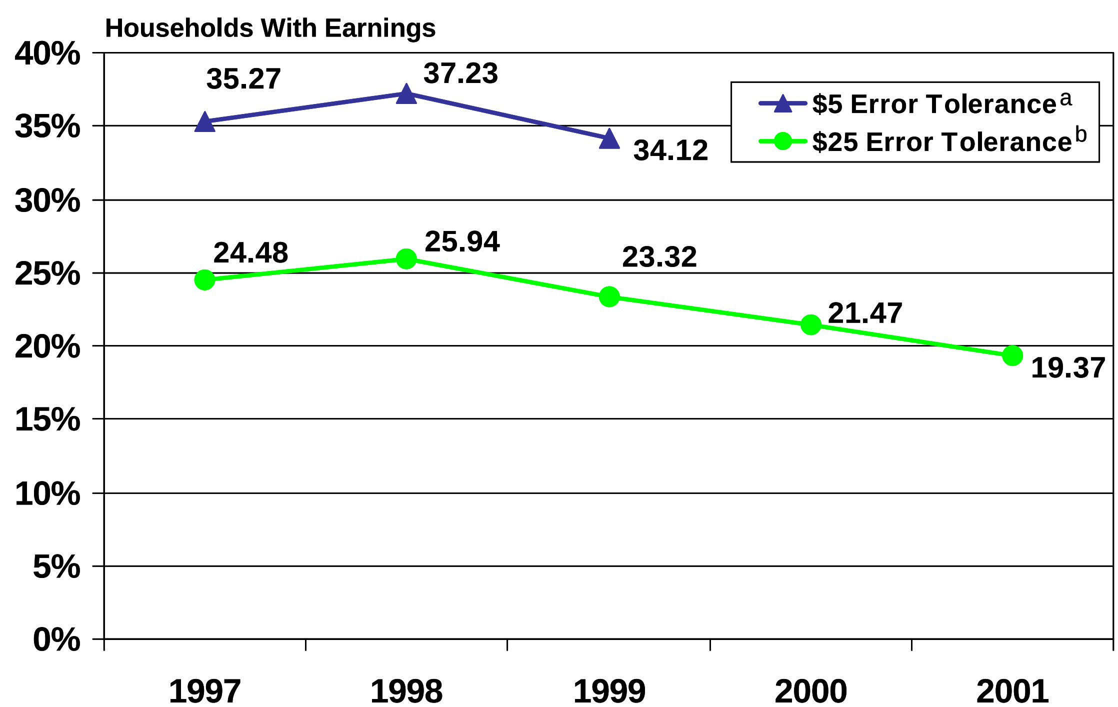
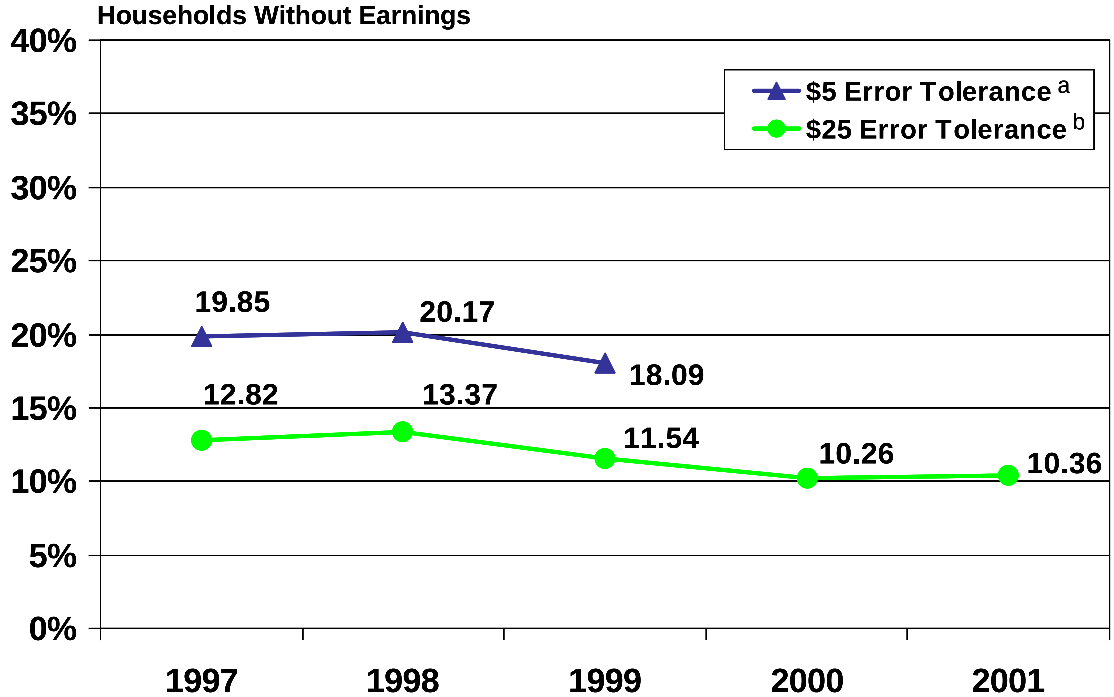


Exhibit 8: National Case Error Rates with $5 and $25 Error Tolerance, 1997-2001 *(Continued)*



Notes:

[a] Eligible cases with overpayment or underpayment of less than $5 are not classified as error cases. (Data not available for 2000 and 2001, as explained in text.)

[b] Eligible cases with overpayment or underpayment of less than $25 are not classified as error cases.

The distinction between cases with earnings and cases without earnings is based on whether the case record indicates the presence of earnings. To some degree, errors among cases "without earnings" are associated with unreported earnings amounts. The model does not address explicitly the extent to which households move between the "with earnings" and "without earnings" categories.

Calculation of Cell Counts and Transition Probabilities

The model's rules for determining cell counts within the five-by-five transition matrix are described in Appendix B. In applying these rules, one uses the information from the QC review to infer the prior-month status of active cases. We tested numerous alternative specifications of these rules, with specific attention to the resulting closure rates. One consideration was to minimize the number of instances of *negative* group-specific closure rates. Such anomalous values were present to some degree under all specifications, reflecting the variability of the weighted sample data as well as possible oversimplification in the modeling assumptions.

We applied the procedures in Appendix B in calculating the cell entries within the five-by-five matrix nationally for 1998 through 2001. For 2001, these estimates are shown in Exhibit 9. (Appendix Exhibits C-1 through C-4 show the estimates for each of the four years). We then used the estimated cell counts to compute the associated transition probabilities for each period. These are shown in Exhibit 10 for 2001 (and in Appendix Exhibits C-5 through C-8 for all four estimated years). For any given year, the transition probabilities in each row sum to 1, as they are computed by dividing each cell count by its corresponding row total. To recall, each transition probability p_{ij} indicates the probability that a household of status i in the current month will occupy status j in the next month.

Derivation of Row Totals

The row totals in the basic transition matrix (R_1 through R_5 in Exhibit 6) represent the current-month (March) distribution of households. ***Row totals cannot be directly estimated from the QC data, because cases that have just closed (the first entry in each row) are not observed in the next-month (April) QC data.*** Some simplifying assumption is thus required to derive these row totals, which can then in turn be used to calculate the number of case closures.

In this section, we discuss at some length the approach used in setting the row totals. The modeling assumptions here deserve attention, for several reasons. First, there are a number of different approaches that one can take, with no compelling logic pointing toward any single correct method. Second, to the extent that different approaches yield different closure rates, the model's estimates will vary according to one's choice of method.

In principle, one would ideally want the marginal totals in the basic transition matrix to reflect both (a) the change from one month to the next in the total population of households and (b) the shift from one month to the next in the distribution of households among the five groups, within the total population count. The approach adopted here is to derive the row totals in such a way as to incorporate the month-to-month distributional shift, as described below, but not the month-to-month population change. The assumption of a constant total population within a given fiscal year is only a very slight abstraction from reality, as the monthly trend rate of national population growth during this period was 0.1 percent.

The model assumes that, over the course of a fiscal year, the change in the distribution of households takes place in twelve equal monthly steps. ***The month-to-month change in the distribution of households across the five groups is computed as one-twelfth of the observed year-to-year change in the household distribution***.

Exhibit 9: Cell Counts, 2001

Current-month ("March") status	Next-month ("April") status: Non-participating	Ongoing correct	Ongoing error	Expiring correct	Expiring error	Total (row)
	Total households (in thousands)					
Nonparticipating	98,385	574	57	0	0	99,015
Ongoing correct	485	4,302	337	633	56	5,814
Ongoing error	23	380	360	2	63	829
Expiring correct	72	509	58	0	0	640
Expiring error	51	56	14	0	0	121
Total (column)	99,015	5,821	827	636	119	106,418
	Households with earnings (in thousands)					
Nonparticipating	83,049	190	29	0	0	83,268
Ongoing correct	141	846	160	211	29	1,387
Ongoing error	43	159	102	1	27	331
Expiring correct	10	169	33	0	0	212
Expiring error	25	28	4	0	0	58
Total (column)	83,268	1,391	329	212	56	85,257
	Households without earnings (in thousands)					
Nonparticipating	15,318	384	27	0	0	15,729
Ongoing correct	344	3,456	177	422	27	4,427
Ongoing error	0*	221	259	1	36	517
Expiring correct	60	340	25	0	0	425
Expiring error	25	28	10	0	0	63
Total (column)	15,747	4,430	498	423	63	21,161

Note: Row (or column) entries may not sum to the indicated row (or column) total due to rounding. Asterisk (*) indicates an imputed zero value.

Exhibit 10: Transition Probabilities, 2001

Current-month (“March”) status	Next-month (“April”) status: Non-participating	Ongoing correct	Ongoing error	Expiring correct	Expiring error	Total
	Total households					
Nonparticipating	0.994	0.006	0.001	0.000	0.000	1.000
Ongoing correct	0.083	0.740	0.058	0.109	0.010	1.000
Ongoing error	0.028	0.459	0.435	0.003	0.076	1.000
Expiring correct	0.113	0.796	0.091	0.000	0.000	1.000
Expiring error	0.420	0.462	0.119	0.000	0.000	1.000
	Households with earnings					
Nonparticipating	0.997	0.002	0.000	0.000	0.000	1.000
Ongoing correct	0.102	0.610	0.116	0.152	0.021	1.000
Ongoing error	0.128	0.480	0.307	0.004	0.081	1.000
Expiring correct	0.049	0.795	0.157	0.000	0.000	1.000
Expiring error	0.440	0.484	0.076	0.000	0.000	1.000
	Households without earnings					
Nonparticipating	0.974	0.024	0.002	0.000	0.000	1.000
Ongoing correct	0.078	0.781	0.040	0.095	0.006	1.000
Ongoing error	0.000*	0.445	0.520	0.002	0.072	1.039
Expiring correct	0.141	0.800	0.059	0.000	0.000	1.000
Expiring error	0.395	0.447	0.158	0.000	0.000	1.000

Note: Row entries may not sum to the indicated row total due to rounding. Asterisk (*) indicates an imputed zero value.

To illustrate the linear interpolation method used to calculate the row totals for the annual national estimates, we describe below how we calculated the national row totals for the 2001 matrix; these are the marginal row totals shown in the top panel of Exhibit 9. As previously noted, we refer to these row totals as the "March 2001" estimates, as they represent the distribution of households one month before the "April 2001" column totals.

Columns (a) through (f) of Exhibit 11 show the calculations for this example. In this exhibit, the weighted household counts are not rounded. The corresponding numbers in Exhibit 8, as elsewhere in the report, are rounded to the nearest thousand.

The steps in the calculation of the March 2001 row totals are as follows:

- Start with the household counts determined in the prior year, 2000. These "April 2000" counts (by group) are the column totals from the top panel of Exhibit C-3, as repeated in column (a) of Exhibit 11. For the four subgroups of active cases, these counts are obtained directly from the 2000 QC data. The total household count for April 2000 (104.705 million, from Census data) is used to derive the April 2000 count of nonparticipating households, as a residual.

- Compute the 2000 ("April 2000") percentage distribution of households across the five groups. This is shown in column (b) of Exhibit 11.

- Apply the percentage distribution of households in column (b) to the national population base for April 2001 (106.418 million, from Census data), thus deriving the counts for each group that would result if the percentage distribution of households had remained unchanged from 2000 to 2001 (i.e., from April 2000 to April 2001). These counts are shown in column (c) of Exhibit 11. This step is in keeping with the model's focus on the aggregate participation rate (versus the number of active cases) and on the case error rate (versus the number of error cases).

- Column (d) shows the household counts for 2001 ("April 2001"), as shown in the column totals in the top panel of Exhibit 9. Compute the average monthly change for each group as 1/12 of the difference between the "April 2001" household count in column (d) and the population-adjusted "April 2000" household count in column (c). For each group, this average monthly change is shown in column (e). It can be either a positive or negative value. By definition, the average monthly changes for the five groups sum to zero.

Exhibit 11: Calculating "March 2001" Row Totals for FY 2001 Transition Matrix

Household group	FY2000 ("April 2000") count	FY2000 ("April 2000") distribution	FY2000 distribution, applied to FY2001 population	FY2001 ("April 2001") count	Average monthly change, FY2000 to FY2001	FY2001 ("March 2001") row total
	(a)	(b)	(c)	(d)	(e)	(f)
Nonparticipating	97,415,884	93.04%	99,009,632	99,015,357	477	**99,014,880**
Ongoing correct	5,642,513	5.39%	5,734,826	5,821,020	7,183	**5,813,837**
Ongoing error	832,461	0.80%	846,080	827,098	-1,582	**828,680**
Expiring correct	676,094	0.65%	687,155	635,508	-4,304	**639,812**
Expiring error	138,048	0.13%	140,306	119,017	-1,774	**120,791**
Total	104,705,000	100.00%	106,418,000	106,418,000	0	106,418,000
	See column totals in top panel of Exhibit C-3	=(a)/104,705,000	=(b)*106,418,000	See column totals in top panel of Exhibit C-4	=[(d)-(c)]/12	=(d)-(e) See row totals in top panel of Exhibit C-4

- For each group, compute the marginal row total for 2001 ("March 2001") by subtracting the average monthly change from the 2001 ("April 2001") count. The result of this calculation is shown in column (f). These counts, rounded to the nearest thousand, then appear as the marginal row totals in the top panel of Exhibit 9.[17]

The marginal row totals provide the basis for calculating the number of monthly case closures for each of the defined subgroups of active cases (ongoing correct, ongoing error, expiring correct, and expiring error). Case closures are not observed in the active case QC data and also cannot be derived by group from negative action QC data. The marginal row totals allow one to compute the number of closures as the row residual. These closure counts comprise the first column of the basic transition matrix. Although the cell counts elsewhere in the five-by-five matrix are not contingent on the marginal row totals, the row totals form the denominators for computing the transition probabilities in each row.

Long-Term Outcomes Implied by the Transition Matrix

One advantage of the Markov modeling approach is that the estimated transition matrix can be used to compute the projected long-term distribution of households across the specified groups. This projected steady-state distribution is the system outcome that would ultimately occur if the process of month-to-month transitions were to continue indefinitely. One can expect that shifts in the underlying pattern of participation and error, such as those associated with more frequent recertification, will play themselves out progressively over time. For this reason, it is useful to assess such shifts in terms of the long-run outcomes.

In the previous chapter, the equilibrium value of the aggregate participation rate was derived for the two-group model as $a/(a+b)$, where a is the case opening rate and b is the case closure rate. We show below the formulas for deriving the equilibrium values for the aggregate participation rate (p^*) and the case error rate (r^*) from the five-group model, using the notation introduced in Exhibit 6. These formulas express each of these long-term outcomes as a function of the associated opening rates and closure rates. In turn, these opening rates and closure rates are computed from cell counts contained in the basic transition matrix.

The formula below for the long-term participation rate is a direct extension of the calculation shown earlier for the two-group model. The long-term case error rate is derived below as a ratio of two proportions, both expressed as shares of the total household population. One (e^*) is the long-term share of the population consisting of error cases; the other (p^*) is the long-term participation rate.

---

[17] In this particular example, the computed marginal row total for nonparticipating households, 99.015 million (for "March 2001"), equals the corresponding first-column total, also 99.015 million (for "April 2001"). These two values are not equal by assumption.

Long-term aggregate participation rate $= p^*$
$= a/(a+b)$

Where a = case opening rate
$= (Q_{12} + Q_{13})/R_1$

b = case closure rate
$= (Q_{21} + Q_{31} + Q_{41} + Q_{51})/(R_2 + R_3 + R_4 + R_5)$

Long-term case error rate $= r^*$
$= e^*/p^*$

Where e^* = aggregate error rate
$= c/(c+d)$

c = error opening rate
$= (Q_{13} + Q_{23} + Q_{24} + Q_{43})/(R_1 + R_2 + R_4)$

d = error closure rate
$= (Q_{31} + Q_{32} + Q_{34} + Q_{51} + Q_{52})/(R_3 + R_5)$

Using these formulas, Exhibit 12 shows the long-term participation rate and case error rate implied by the 2001 transition matrices for total households, households with earnings, and households without earnings. For each segment of the population, the matrices imply a slight downward or stable path for both the participation rate and the error rate.[18]

[18] An alternative specification of the model was suggested by the Economic Research Service, employing a different approach to the calculation of row totals in the transition matrix that yielded lower case closure rates. This alternative model was considered but was not adopted, in part because the long-term outcomes implied by the transition matrix appeared less credible than those shown in Exhibit 12. In particular, under the alternative formulation, the long-term participation rate for total households was 7.91 percent, implying a substantial increase in the active caseload. The corresponding long-term case error rate was 11.03 percent.

Exhibit 12: Projected Long-Term Outcomes, Based on 2001 Matrix

	Observed FY 2001 value (%)	Projected long-term value (%)
Total households		
Aggregate participation rate	6.96	6.95
Case error rate	12.78	12.72
Households with earnings		
Aggregate participation rate	2.33	2.33
Case error rate	19.37	19.19
Households without earnings		
Aggregate participation rate	25.58	24.86
Case error rate	10.36	10.28

Source: Based on formulas shown in text, using the transition matrices derived for 2001 (see Exhibit 9).