MORE ON CONTROLLED SELECTION

Steven G. Heeringa and Irene Hess, Survey Research Center, The University of Michigan

INTRODUCTION: Controlled selection is a probability sampling procedure which enables its users to purposively introduce constraints on the distribution of the sample. There is also an expectation of reduced sample variances relative to those from other stratified sample designs.

For more than 30 years, the University of Michigan's Survey Research Center has used this technique in developing samples for research studies. The Bureau of the Census has used controlled selection for the Current Population Survey samples in order to control sample distributions by states (U.S. Bureau of the Census, 1963, 1978). A past revision of the Consumer Price Index sample also applied controlled selection techniques to achieve rigorous geographic control (Wilkerson,1961). Gains in precision resulting from controlled selection have been explored, but not enough is yet known about that aspect of potential benefits from the use of controlled selection.

The purpose of a current research investigation at the Survey Research Center is twofold: (1) To observe the performance of various forms of controlled selection when compared among themselves as well as with other selection modes; (2) To investigate ways to improve the application of controlled selection to sampling methods.

The purpose of this paper is to give a progress report on some of the research undertaken to date.

There are two principal findings: (1) For the same sample design, computer generated controlled selection often leads to slightly higher variances than does manual controlled selection; but since the differences in precision are small and manual controlled selection is laborious, computer generated controlled selection is preferred; (2) With stratified random sampling as a basis for comparison, computer generated controlled selection, and ordered systematic selection both result in lower between primary sampling unit variances, and the larger reduction generally is effected by controlled selection.

DATA USED IN THE RESEARCH: *Research populations*: In order to simulate sampling operations, data from the 1960 census were employed in developing alternative sample designs that were then tested with census data from later periods. Counties or county groups in the North Central and South Regions of the United States were regarded as separate populations.

Excluded from each region are the Standard Metropolitan Statistical Areas (SMSA's) with 1960 populations in excess of 500,000 inhabitants. It is assumed that those areas would be included with certainty in any regional sample. The proportion of 1960 inhabitants remaining in each region after excluding major urban areas was a little over half in the North Central and about three-fourths in the South.

Formation of primary sampling units: The primary sampling units are SMSA's, single counties or groups of geographically contiguous counties combined to meet a minimum size of approximately 20,000 inhabitants in 1960. In the North Central 777 psu's were formed, in the South 1,046.

Stratification: In addition to the separation of SMSA's and non-SMSA's, two other dimensions in stratification were observed: geographic location and level of urbanization. Geographic classifications were states or state groups. Urban categories were developed from the 1960 census reports of percent urban for counties and SMSA's.

Selection probabilities: In order to proceed with the research, it was necessary to specify sample sizes. In the North Central 40 selections are assumed, with exactly 12 from SMSA's and 28 from non-SMSA's. In the South 56 selections are assumed, 18 SMSA's and 38 non-SMSA's. The sizes remain constant for all investigations. Selection probabilities were calculated separately for SMSA's and non-SMSA's of each region.

The choice of variables: The aggregates listed in Table 1 were selected from the 1972 *County City Data Book* [U.S. Bureau of the Census, 1973], with the intent to include varibles describing persons, housing units or households, farms, businesses and industries. Census data reported as percentages were converted to aggregates and regarded as exact values thereafter.

Table 2 contains ratios of some of the aggregates listed in Table 1.

CONTROLLED SELECTION DESIGNS USED IN THE RESEARCH: Different forms of stratification for controlled selection as well as a simple example of the procedure are given in Appendix A. Three controlled selection designs, all of one type, have been used so far in the research. For the type used, selection probabilities sum to an integer over rows and columns but not necessarily by rows or by columns. Designs I and II relate to the North Central Region, III to the South.

Design I is a cross-tabulation of six state groups by 11 urban classes - 7 for non-SMSA's and 4 for SMSA's. The psu probabilities sum to exactly 28 over all non-SMSA's, 12 over all SMSA's. In the matrix of 66 cells, 49 are nonzero.

Design II differs from I by increasing the state categories from six to 12, and the matrix from 66 to 132 cells, of which 86 are nonzero. Design II generally has lower variances--a result of increased stratification.

Design III in the South is similar to the North Central designs. The psu's were assigned to seven state groups and 12 urban categories. There are 69 nonzero cells.

MANUAL CONTROLLED SELECTION COMPARED WITH COMPUTER CONTROLLED SELECTION: Does the application of controlled selection to sampling procedures require the services of an experienced sampler, or can computers be programmed to perform the operations satisfactorily? The displays in Tables 3 and 4 are responses to that question. Both manual and computer controlled selections were completed for each of the three designs. Performance tests are based on the precision of estimates achieved by each process.

The relative variances from manual controlled selections are the bases with which relative variances from computer generated controlled selections are compared. Variable code numbers in Table 3 correspond to those in Table 1 where each variable is described. Similarly, the ratios in Table 4 are described in Table 2.

The variances reported in Tables 3 and 4 are between psu relative variances which are population values. No sampling of primary units occurs.

In columns 3, 5 and 7 of Table 3, a quotient in excess of *one* indicates higher variance and lower precision from computer selections. The average quotients and ranges are given at the bottom of the table. Neither procedure shows a clear advantage.

Table 1. AGGREGATES INCLUDED IN NUMERICAL CALCULATIONS, NORTH CENTRAL AND SOUTH REGIONS OF THE UNITED STATES EXCLUSIVE OF MAJOR METROPOLITAN AREAS

Variable Code Nos.	Variable Descriptions	Regions* North Central (part)	South (part)
1	Population, 1970	31,466,373	45,278,109
2	Civilian labor force	12,232,279	16,616,911
3	Females	4,460,926	6,353,671
4	Married, husband present	2,713,839	3,951,887
5	Persons 65 years and over	3,505,126	4,559,033
6	Persons below low income level in 1969	3,695,015	10,091,864
7	Persons 3-34 years enrolled in college	1,177,522	1,223,663
8	Farm population	3,853,953	2,899,976
9	Persons below low income level in 1969	572,608	784,215
10	Black population	841,593	8,316,842
11	Persons of Spanish heritage	229,368	1,553,119
12	Vote cast for President, 1968	12,451,970	14,137,207
13	Families, 1970	7,911,202	11,508,220
14	With income $25,000 or more in 1969	246,806	282,872
15	Recipients of aid to families with dependent children, Feb. 1972	994,005	2,012,089
16	Public assistance payments, Feb. 1972 ($1,000)	82,015	130,899
17	Year-round housing units, 1970	10,506,469	15,065,608
18	Occupied housing units	9,730,068	13,760,894
19	With telephone available	8,798,148	10,771,490
20	With home food freezer	3,974,586	4,895,629
21	Household head moved into unit during 1965-70	4,464,355	6,853,436
22	One person households	1,677,300	2,109,300
23	Farms, 1969 Census of Agriculture	1,113,510	1,129,186
24	With sales $2,500 and over	844,926	566,972
25	Value of farm products sold by farms with sales of $2,500 and over, 1969 ($1,000)	19,238,295	12,387,695
26	Value of livestock and livestock products sold, 1969 ($1,000)	10,084,211	4,041,223
27	Manufacturing estab's, 1967 Census of Manufactures	42,390	56,826
28	Establishments with 100 or more employees	5,263	7,395
29	Mineral industries establishments, 1967 Census of Mineral Industries**	8,612	24,020
30	Retail trade establishments 1967 Census of Business	314,083	427,658
31	Sales of estab's ($1,000)	48,863,904	58,558,573
32	Estab's with payroll	221,488	271,097

* Entries in these columns will not agree with figures shown in Census Bureau publications, for three reasons: 1) major metropolitan areas have been excluded; 2) to avoid dividing SMSA's that crossed regional boundaries, a clear separation of regions was sacrificed; 3) where data of interest to the research were reported as percents in the *County and City Data Book*, they were converted to aggregates and regarded thereafter as exact values.
** Mineral industries establishments shown in this table are sums of county totals, which do not agree with state totals reported in the ***County and City Data Book.***
Source: U.S. Bureau of the Census, 1973.

Turning to Table 4, consider the performance of each selection design when estimates are ratios. Averages of quotients, given below the table, are a little more than *one*, indicating some increased variance on the average from computer generated controlled selection.

Over all observations, manual controlled selection appears to give slightly higher precision within the bounds of the tests. Nevertheless, the authors agree that the laborious process of manual controlled selection has hindered its

Table 2. RATIOS INCLUDED IN NUMERICAL CALCULATIONS, NORTH CENTRAL AND SOUTH REGIONS OF THE UNITED STATES EXCLUSIVE OF MAJOR METROPOLITAN AREAS

Ratio Code Nos.	Ratio Descriptions	Num. & Denom. Code Nos.*	Regions ** North Central (part)	South (part)
40	Percent of total population voting for President, 1968	12,1	39.572	31.223
41	Percent Black population	10,1	2.675	18.368
42	Percent of persons 65 years of age and over	5,1	11.139	10.069
43	Percent Spanish heritage population	11,1	0.729	3.430
44	Percent persons 3-34 years enrolled in college	7,1	3.742	2.703
45	Percent females in civilian labor force	3,2	36.468	38.236
46	Percent females in labor force married with husband present	4,3	60.836	62.198
47	Percent families with income $25,000 or more	14,13	3.120	2.458
48	Average public assistance payment per family	16,13	$10.367	$11.374
49	Percent farm population below low income level	9,8	14.858	27.042
50	Percent occupied housing units	18,17	92.610	91.340
51	Percent occupancy by one-person households	22,18	17.238	15.328
52	Percent occupancy by movers into units during 1965-1970	21,18	45.882	49.804
53	Percent occupancy with home freezer	20,18	40.848	35.576
54	Percent occupancy with telephone available	19,18	90.422	78.276
55	Percent farms with sales $2,500 or more	24,23	75.880	50.211
56	Livestock and livestock products sales as percent of total sales by farms with sales of $2,500 or more	26,25	52.417	32.623
57	Percent manufacturing establishments with 100 or more employees	28,27	12.416	13.013
58	Average sales per retail trade establishment ($1,000)	31,30	155.576	136.929
59	Percent retail trade establishments with payroll	32,30	70.519	63.391

* See Table 1 for identification.
** See footnotes to Table 1.

acceptance and probably would continue to do so, whereas controlled selection by computers gives acceptable precision and places the sampling technique on a practical basis.

CONTROLLED SELECTION COMPARED WITH STRATIFIED RANDOM SAMPLING AND ORDERED SYSTEMATIC SAMPLING: How does the precision of estimates from controlled selection compare with that from other selection methods? Comparisons are made with two other sampling procedures: stratified random sampling, and ordered systematic sampling, each method designed to have exactly two selections per stratum. These two methods differ only in the way primary units are paired within strata. The procedures are discussed and illustrated in Appendix B.

Table 3. BETWEEN PSU RELATIVE VARIANCES OF AGGREGATES FOR THREE MANUAL CONTROLLED SELECTION DESIGNS, AND COMPARISONS WITH RELATIVE VARIANCES FOR COMPUTER GENERATED CONTROLLED SELECTIONS

Variable	North Central				South	
	Design I†		Design II†		Design III†	
Code Nos.*	$V^2_{y_{cs-man}}$ ‡	$V^2_{y_{cs-comp}}$ § / $V^2_{y_{cs-man}}$ ‡	$V^2_{y_{cs-man}}$ ‡	$V^2_{y_{cs-comp}}$ § / $V^2_{y_{cs-man}}$ ‡	$V^2_{y_{cs-man}}$ ‡	$V^2_{y_{cs-comp}}$ § / $V^2_{y_{cs-man}}$ ‡
1	2	3	4	5	6	7
1	.000350	1.003	.000359	.966	.000525	1.009
2	.000519	1.002	.000559	.887	.000754	.987
3	.000777	.991	.000851	.861	.000866	.981
4	.000760	.992	.000813	.895	.001109	.992
5	.000822	.995	.000748	1.052	.002789	1.028
6	.002309	1.008	.002017	.986	.001414	.996
7	.041173	.985	.041290	.951	.028098	1.005
8	.004744	1.031	.004208	1.007	.010901	1.004
9	.014536	1.015	.013022	.994	.022607	.999
10	.020330	1.091	.018800	1.109	.006271	.989
11	.038245	1.064	.037380	1.031	.123861	1.006
12	.000353	.986	.000360	.983	.000849	1.003
13	.000305	.999	.000317	.977	.000649	1.013
14	.003669	1.006	.003607	.896	.005025	1.012
15	.017069	1.013	.015270	1.009	.003780	.996
16	.012085	1.000	.009511	1.002	.003144	1.034
17	.000439	.992	.000422	1.033	.000714	1.027
18	.000288	.997	.000302	.960	.000658	1.020
19	.000341	1.011	.000355	.966	.000928	1.027
20	.000799	1.007	.000793	.986	.001120	1.016
21	.000922	1.014	.000905	.950	.001486	1.012
22	.000621	.934	.000560	1.031	.001238	1.048
23	.004278	1.021	.004109	.973	.007984	1.009
24	.005736	1.034	.004944	1.022	.012578	1.026
25	.013664	1.028	.010298	1.023	.046291	1.012
26	.029450	1.009	.022123	1.006	.206458	1.005
27	.003697	.970	.003362	.971	.004404	1.013
28	.007021	.986	.006523	.990	.007580	1.002
29	.169167	.995	.148758	1.002	.071519	1.002
30	.000779	1.031	.000758	.981	.000662	1.005
31	.000710	1.057	.000653	1.055	.000929	.956
32	.000747	1.015	.000721	.991	.000751	1.009
Summary Measures						
Arithmetic means		1.009	NA	.986	NA	1.008
Ranges		.970-1.091	NA	.861-1.109	NA	.956-1.048

NA Not applicable. * See Table 1 for variable descriptions.
† See Text for design descriptions. Design I has 40 selections from 49 cells; Design II has 40 selections from 86 cells; Design III has 56 selections from 69 cells.
‡ $V^2_{y_{cs-man}}$ denotes relative variances for manually constructed controlled selection.
§ $V^2_{y_{cs-comp}}$ denotes relative variances for computer generated controlled selections.

The stratified designs use the same primary units, the same stratification variables, the same number of sample selections, and with few exceptions the same psu probabilities as were used for the controlled selection designs. Occasionally, psu probabilities were adjusted slightly so that their sums within strata would be exactly *two*. No psu crosses a stratum boundary. The stratified designs share the same strata, which were constructed to satisfy the ordered systematic design.

For the randomized design, psu's within strata were rearranged in a random order as described in Appendix B.

The ordered systematic design may be regarded as a form of controlled selection, or, as Goodman and Kish discussed in their 1950 paper, controlled selection may be viewed as systematic sampling when the primary units are ordered in a meaningful sequence.

Table 4. BETWEEN PSU RELATIVE VARIANCES OF RATIOS FOR THREE MANUAL CONTROLLED SELECTION DESIGNS, AND COMPARISONS WITH RELATIVE VARIANCES FROM COMPUTER GENERATED CONTROLLED SELECTIONS

Variable	North Central				South	
	Design I†		Design II†		Design III†	
Code Nos.*	$V^2_{y_{cs-man}}$ ‡	$V^2_{y_{cs-comp}}$ § / $V^2_{y_{cs-man}}$ ‡	$V^2_{y_{cs-man}}$ ‡	$V^2_{y_{cs-comp}}$ § / $V^2_{y_{cs-man}}$ ‡	$V^2_{y_{cs-man}}$ ‡	$V^2_{y_{cs-comp}}$ § / $V^2_{y_{cs-man}}$ ‡
1	2	3	4	5	6	7
40	.000300	.990	.000267	1.037	.000560	1.014
41	.020502	1.097	.018991	1.112	.006952	1.007
42	.001101	.982	.000987	1.030	.002342	1.016
43	.037387	1.064	.036438	1.042	.123578	1.006
44	.038384	.984	.038454	.956	.026443	1.001
45	.000099	.970	.000102	.888	.000090	1.051
46	.000094	1.000	.000091	.981	.000072	1.034
47	.003011	1.008	.002867	.921	.003305	1.015
48	.012347	1.000	.009763	1.004	.004075	1.028
49	.007202	1.018	.006417	.999	.007405	1.008
50	.000096	1.001	.000089	1.038	.000020	1.018
51	.000411	.927	.000358	1.029	.000499	1.060
52	.000387	1.028	.000366	.988	.000342	1.008
53	.000551	.988	.000522	.932	.000917	1.060
54	.000040	1.030	.000032	1.099	.000089	1.026
55	.000886	1.019	.000533	1.115	.003410	1.024
56	.006269	.996	.005087	.999	.080078	1.001
57	.004157	.998	.003835	1.096	.006007	.998
58	.000742	1.022	.000688	.956	.000533	.961
59	.000106	1.031	.000084	1.059	.000108	1.019
Summary Measures						
Arithmetic means		1.008	NA	1.014	NA	1.018
Ranges		.927-1.097	NA	.888-1.115	NA	.961-1.060

NA Not applicable. * See Table 2 for ratio descriptions.
† See Text for design descriptions. Design I has 40 selections from 49 cells; Design II has 40 selections from 86 cells; Design III has 56 selections from 69 cells.
‡ $V^2_{y_{cs-man}}$ denotes relative variances for manually constructed controlled selection.
§ $V^2_{y_{cs-comp}}$ denotes relative variances for computer generated controlled selections.

With relative variances from stratified random sampling as a base, in Tables 5 and 6 ordered systematic sampling and controlled selection are compared to the randomized design and then with each other. The construction of Tables 5 and 6 parallels that of Tables 3 and 4. In the North Central Region, only Design II of the three computer controlled selection designs is included. Table 5 contains the comparisons for aggregates. The results for ratios are given in Table 6. The variances are between psu relative variances, which are population values. No sampling occurs. The summary measures below the tables show that all column averages are below *one*, indicating variance reductions. Notice that the lowest figures are associated with controlled selection. The ranges related to controlled selection are wider but have lower bounds than those for ordered systematic selection. Outcomes might be different if research conditions were changed. But for the present investigation, controlled selection shows an advantage.

However, a few *words of warning* are in order. The observed relative variances are between psu variances. What the effect on total variance might be we are unable to say at this time. The within-psu component can vary widely for different populations and with different characteristics of the same population. Furthermore, when psu's selected with certainty are combined with non-certainty selections, the effect of the between psu variance component will be reduced.

COMPONENTS OF CONTROLLED SELECTION VARIANCES: The opportunity to study between and within pattern variance components is a dividend from controlled selection variance calculations as illustrated in Table 7 for selected ratios from manual and computer selections, Designs I and II in the North Central.

In Table 7, the pairs of variance components from the two designs illustrate that, for a given sample design and ratio, the within pattern variance components are constant no matter how many patterns are formed or by what means. *Also notice that*: (1) The within pattern component dominates the total variance; (2) A reduction in within pattern variances and an increase in the between components is the general result from the increased stratification in Design II. But the net result is a reduction in total variance.

Some optimal balance of within and between pattern variance components is needed to reduce both the within pattern components and the total variances. This is an activity for continuing research.

ESTIMATING SAMPLING VARIABILITY FROM A CONTROLLED SELECTION SAMPLE: There has been some reluctance to use controlled selection because it is not a measurable design. To obtain approximate measures of sampling variability from a single sample researchers may choose a technique designed for systematic sampling, a reasonable choice to make as the two sampling procedures are closely related.

Table 5. BETWEEN PSU RELATIVE VARIANCES OF AGGREGATES FOR STRATIFIED RANDOM SELECTIONS, COMPARED WITH RELATIVE VARIANCES FROM ORDERED SYSTEMATIC SELECTIONS AND FROM COMPUTER GENERATED CONTROLLED SELECTIONS

Variable Code Nos.*	North Central: $V^2_{y_{sr}}$ † (20 strata, 40 sels.)	North Central: $V^2_{y_{sys}}$ ‡ / $V^2_{y_{sr}}$ †	North Central: $V^2_{y_{cs-comp}}$ § / $V^2_{y_{sr}}$ †	North Central: $V^2_{y_{cs-comp}}$ § / $V^2_{y_{sys}}$ ‡	South: $V^2_{y_{sr}}$ † (28 strata, 56 sels.)	South: $V^2_{y_{sys}}$ ‡ / $V^2_{y_{sr}}$ †	South: $V^2_{y_{cs-comp}}$ § / $V^2_{y_{sr}}$ †	South: $V^2_{y_{cs-comp}}$ § / $V^2_{y_{sys}}$ ‡
1	.000371	.972	.935	.962	.000558	.905	.949	1.049
2	.000520	.968	.953	.984	.000759	.901	.980	1.087
3	.000766	1.027	.957	.932	.000940	.860	.905	1.052
4	.000776	1.004	.937	.933	.001228	.885	.896	1.013
5	.000928	.972	.849	.873	.003462	.824	.828	1.005
6	.002333	.912	.852	.935	.001707	.916	.825	.901
7	.042257	.973	.929	.955	.028347	.933	.997	1.068
8	.005042	.937	.840	.897	.011726	.878	.933	1.062
9	.014787	.930	.875	.941	.023536	.912	.960	1.053
10	.024203	1.001	.862	.861	.008975	.958	.691	.721
11	.042287	.926	.912	.984	.143963	.818	.865	1.057
12	.000413	.911	.857	.941	.000898	.918	.948	1.033
13	.000318	.920	.975	1.060	.000681	.912	.965	1.058
14	.003631	.973	.890	.915	.005724	.876	.889	1.014
15	.017755	.993	.868	.874	.005041	.897	.747	.833
16	.013538	.982	.704	.717	.003834	.944	.848	.899
17	.000449	.935	.971	1.039	.000774	.901	.947	1.052
18	.000298	.945	.971	1.028	.000702	.898	.957	1.066
19	.000344	.955	.996	1.043	.000992	.906	.960	1.060
20	.000866	.934	.904	.967	.001218	.918	.934	1.018
21	.000949	1.065	.906	.851	.001678	.871	.896	1.029
22	.000685	.947	.844	.890	.001470	.849	.882	1.039
23	.004228	.957	.945	.988	.009315	.889	.865	.973
24	.006346	.939	.796	.848	.014861	.917	.868	.947
25	.015520	.972	.679	.698	.048236	.961	.971	1.010
26	.034428	.988	.646	.655	.223552	.997	.928	.932
27	.003786	.899	.863	.959	.004657	1.017	.958	.942
28	.006711	.920	.962	1.045	.008695	.955	.873	.914
29	.199799	1.025	.746	.728	.077930	.965	.920	.953
30	.000962	.931	.773	.830	.000743	.935	.896	.958
31	.000749	.933	.920	.986	.000969	.905	.917	1.014
32	.000859	.914	.832	.910	.000866	.947	.876	.924
Summary Measures								
Arithmetic means		.958	.873	.913	NA	.912	.902	.992
Ranges		.899-1.065	.646-.996	.655-1.060	NA	.824-1.017	.691-.997	.721-1.087

NA Not applicable. * See Table 1 for variable descriptions.

† $V^2_{y_{sr}}$ denotes relative variances for stratified random selections, two per stratum.

‡ $V^2_{y_{sys}}$ denotes relative variances for ordered, systematic selections, two per stratum (20 strata in the North Central; 28 strata in the South).

§ $V^2_{y_{cs-comp}}$ denotes relative variances for computer generated controlled selections (40 selections from 86 cells in the North Central; 56 selections from 69 cells in the South).

Table 6. BETWEEN PSU RELATIVE VARIANCES OF RATIOS FOR STRATIFIED RANDOM SELECTIONS, COMPARED WITH RELATIVE VARIANCES FROM ORDERED SYSTEMATIC SELECTIONS AND FROM COMPUTER GENERATED CONTROLLED SELECTIONS

Variable Code Nos.*	North Central: $V^2_{y_{sr}}$ † (20 strata, 40 sels.)	North Central: $V^2_{y_{sys}}$ ‡ / $V^2_{y_{sr}}$ †	North Central: $V^2_{y_{cs-comp}}$ § / $V^2_{y_{sr}}$ †	North Central: $V^2_{y_{cs-comp}}$ § / $V^2_{y_{sys}}$ ‡	South: $V^2_{y_{sr}}$ † (28 strata, 56 sels.)	South: $V^2_{y_{sys}}$ ‡ / $V^2_{y_{sr}}$ †	South: $V^2_{y_{cs-comp}}$ § / $V^2_{y_{sr}}$ †	South: $V^2_{y_{cs-comp}}$ § / $V^2_{y_{sys}}$ ‡
40	.000378	1.007	.733	.728	.000663	.851	.856	1.006
41	.024277	.996	.870	.873	.009510	.969	.736	.759
42	.001234	1.024	.824	.804	.002965	.819	.803	.980
43	.041759	.921	.909	.988	.143591	.820	.866	1.056
44	.039430	.964	.932	.967	.026825	.946	.987	1.043
45	.000098	.966	.924	.956	.000120	.858	.789	.920
46	.000099	.910	.904	.993	.000079	.959	.937	.978
47	.002969	.983	.889	.905	.003735	.874	.898	1.028
48	.013450	.981	.729	.743	.004690	.967	.893	.924
49	.007492	.940	.856	.910	.008289	.945	.900	.953
50	.000098	.922	.949	1.029	.000023	.965	.915	.948
51	.000458	.955	.803	.841	.000616	.856	.860	1.006
52	.000398	1.106	.909	.822	.000393	.854	.877	1.027
53	.000647	.951	.752	.790	.001108	.879	.877	.998
54	.000042	1.017	.849	.835	.000112	.897	.818	.912
55	.001040	.920	.571	.621	.003584	.935	.974	1.042
56	.007472	1.042	.678	.651	.088799	1.006	.903	.898
57	.004415	.955	.952	.998	.006846	.993	.876	.882
58	.000843	.950	.780	.882	.000587	.883	.874	.990
59	.000122	1.015	.726	.715	.000123	.920	.901	.979
Summary Measures								
Arithmetic means		.976	.827	.850	NA	.910	.877	.966
Ranges		.910-1.106	.571-.952	.621-1.029	NA	.819-1.006	.736-.987	.759-1.056

NA Not applicable. * See Table 2 for ratio descriptions.

† $V^2_{y_{sr}}$ denotes relative variances for stratified random selections, two per stratum.

‡ $V^2_{y_{sys}}$ denotes relative variances for ordered, systematic selections, two per stratum (20 strata in the North Central; 28 strata in the South).

§ $V^2_{y_{cs-comp}}$ denotes relative variances for computer generated controlled selections (40 selections from 86 cells in the North Central; 56 selections from 69 cells in the South).

Some investigations have been undertaken on this topic, but accomplishments are not yet sufficient for reporting.

ESTIMATES FOR DOMAINS: Another area for investigation is the precision of estimates of domains that are planned and included in controlled selection designs. Some researchers have suggested that controlled selection might be especially well suited for domain estimation. This suggestion is yet to be tested.

Table 7. COMPONENTS OF CONTROLLED SELECTION VARIANCES FOR SELECTED RATIOS, NORTH CENTRAL REGION

Items	Design I		Design II	
	41 manually constructed patterns	29 computer constructed patterns	47 manually constructed patterns	34 computer constructed patterns
1	2	3	4	5
42. Proportion of persons 65 years of age and over				
Total				
Var. (Y)*	10,100	10,048	9,193	9,674
Var. (X)*	346,957	347,977	355,828	343,793
Covar. (Y,X)*	3,925	4,861	6,648	6,477
Between patterns				
Var. (Y)*	294	242	550	1,031
Var. (X)*	8,453	9,473	42,819	30,783
Covar. (Y,X)*	- 1,023	- 87	- 1,528	- 1,699
Within patterns				
Var. (Y)*	9,806	9,806	8,643	8,643
Var. (X)*	338,504	338,504	313,010	313,010
Covar. (Y,X)*	4,948	4,948	8,176	8,176
52. Proportion occupancy by movers into housing units during 1965-70				
Total				
Var. (Y)*	18,372	18,625	18,031	17,130
Var. (X)*	27,254	27,182	28,575	27,420
Covar. (Y,X)*	17,868	17,889	18,248	17,099
Between patterns				
Var. (Y)*	834	1,087	3,519	2,617
Var. (X)*	671	599	3,504	2,349
Covar. (Y,X)*	638	659	3,207	2,058
Within patterns				
Var. (Y)*	17,539	17,539	14,513	14,513
Var. (X)*	26,583	26,583	25,071	25,071
Covar. (Y,X)*	17,229	17,229	15,041	15,041
58. Average sales per retail trade establishment ($1,000)				
Total				
Var. (Y)*	1,695,422	1,792,213	1,559,891	1,645,162
Var. (X)*	77	79	75	73
Covar. (Y,X)*	5,737	6,113	5,551	5,943
Between patterns				
Var. (Y)*	70,974	167,765	153,427	238,698
Var. (X)*	1	4	7	6
Covar. (Y,X)*	100	476	32	424
Within patterns				
Var. (Y)*	1,624,448	1,624,448	1,406,464	1,406,464
Var. (X)*	75	75	68	68
Covar. (Y,X)*	5,637	5,637	5,519	5,519

* In millions.

Appendix A

CONTROLLED SELECTION ILLUSTRATED: Simple examples of stratification for controlled selection are given in Illustration 1. In each example a population of 18 sampling units is assumed to have been distributed to 12 cells resulting from cross-tabulations by two variables, one with three and the other with four categories. While for simplification the numbers of variables and categories are kept small, each can be increased in practice. Also it is assumed that about 6 sample selections are to be made. The three examples illustrate design variations that achieve the desired sample size while distributing the sample across cells and marginals in proportion to their respective expectations.

In *Example X*, controlled selection can achieve a sample size of 5 with probability .3, or a sample of 6 with probability .7. Similar statements could be made for each cell.

In *Example Y*, selection probabilities for the 18 sample units have been calculated to sum to exactly 6. Therefore, controlled selection would always yield a sample of 6, with each of the marginals and cells tending to contribute proportionately to the total sample according to the indicated probabilities. Row 1 would have 2 selections with probability .6 or 3 selections with probability .4, and so on.

In *Example Z*, selection probabilities for individual units have been adjusted so that each row adds to exactly 2. Adjustments could have been made to other integers, and in practice they are. Controlled selection would now achieve a sample of 6 with 2 selections from each row. Cells and columns would share proportionately in the sample according to their respective expectations.

Illustration 2 displays a set of patterns or samples resulting from the application of controlled selection to the population described by Example Y. Pattern weights or values sum to *1.0*. Across all patterns the selection probabilities or expectations are satisfied exactly for every cell and marginal. Every pattern has *6* selections. Random choice of a pattern provides a sample that satisfies the specifications in Example Y.

Notice that controlled selection does not designate a particular set of sampling units. Controlled selection specifies the number of sampling units to be selected from designated cells. Illustrations 1 and 2 show that multiple selections from cells can occur and are permissible. The variance calculations assume that sampling units are chosen within cells in proportion to assigned probabilities and with replacement when multiple selections are to be made from a cell.

Illustration 1. Three Ways to Stratify Sampling Units in Preparation for Controlled Selection

EXAMPLE X

Groups	Classes				
	A	B	C	D	Total
1	.2	.7	.0	1.3	2.2
2	.5	.3	.4	.2	1.4
3	1.3	.2	.6	.0	2.1
Total	2.0	1.2	1.0	1.5	5.7

EXAMPLE Y

Groups	Classes				
	A	B	C	D	Total
1	.2	.7	.0	1.5	2.4
2	.5	.3	.4	.2	1.4
3	1.4	.2	.6	.0	2.2
Total	2.1	1.2	1.0	1.7	6.0

EXAMPLE Z

Groups	Classes				
	A	B	C	D	Total
1	.2	.6	.0	1.2	2.0*
2	.7	.4	.6	.3	2.0*
3	1.2	.2	.6	.0	2.0*
Total	2.1	1.2	1.2	1.5	6.0

* The number of rows and row totals may be varied so long as rows add to an integer satisfying research needs (e.g., two rows totalling 3; six rows totaling 1).

Illustration 2. Controlled Selection Patterns for Example Y Displayed in Illustration 1

Cell Identification	Initial Probabilities	Pattern Numbers								
		1	2	3	4	5	6	7	8	9
		Pattern Weights*								
		.1	.2	.1	.1	.1	.1	.1	.1	.1
1 A	.2				X		X			
B	.7	X	X			X	X	X	X	
D	1.0	X	X	X	X	X	X	X	X	X
D	.5			X	X	X		X		X
2 A	.5	X	X	X		X				
B	.3		X		X					
C	.4	X		X				X		X
D	.2						X		X	
3 A	1.C	X	X	X	X	X	X	X	X	X
A	.4	X						X	X	X
B	.2			X						X
C	.6		X		X	X	X		X	

* Pattern weights or values must add to 1.0.

Appendix B

ILLUSTRATION OF STRATIFIED DESIGNS WITH TWO SELECTIONS PER STRATUM: Two stratified designs are used in the research: stratified random, and ordered systematic. Simple examples of the selection designs are given in Illustrations 4 and 5, each derived from the strata in Illustration 3, which is an adaptation of Example Y in Illustration 1, Appendix A.

The 18 sampling units are listed separately in a prescribed order that retains stratification by groups 1, 2 and 3, while reversing the order of classes A through D as the listing continues from one group to the next. The last class in group 1 is followed by the same class in group 2, and so on. Some of the unit probabilities were adjusted to force probabilities to total 2 within each stratum. (While adjustments may appear to be major among the 18 primary units, in practice adjustments usually are within rounding error and affect only a few primary units.) With probabilities calculated to one decimal, exactly ten pairs of primary units can be formed within a stratum, each pair receiving a weight of 0.1. (If psu probabilities were calculated to two or three decimals, 100 or 1,000 pairs would be formed in each stratum.) The method used to form pairs of units is the feature that distinguishes stratified random selection from the ordered systematic selection.

The *stratified random selections* displayed in Illustration 4 were made independently within each stratum by drawing 20 numbers at random to fill the ten cells. Numbers 01 and 11 were assigned to cell 1, 02 and 12 to cell 2, and so on. Notice that psu D1 in stratum I had an adjusted probability of .4 in column 8, Illustration 3. Therefore, psu D1 appears four times (in positions 01, 04, 13 and 19) in stratum I, Illustration 4. Also notice that psu B2 with adjusted probability of .4 was assigned to positions 05, 06, 12 and 15. That is, in one cell B2 was paired with itself. Other self-pairings occurred in strata II and III. Utilizing a psu (in the selection process) according to its adjusted probability is inconsistent with random sampling with replacement, which could result in selecting a primary unit more or less frequently than its adjusted probability would indicate. However, the procedure that was used was preferred for the purpose of calculating population variances at a later stage in the research.

Ordered systematic selection restricts the combinations of sampling units that enter a particular sample. Judicious ordering of strata and sampling units are important steps in ordered systematic selection. The arrangement of strata and units in Illustration 3 was designed to satisfy the systematic design, and that ordering was maintained when forming the pairs of psu's shown in Illustration 5.

Notice the difference between the assignment of primary untis to cells in Illustrations 4 and 5. In stratum I of Illustration 5, psu D1 is assigned to four positions in sequence, 01, 02, 03, and 04. Then psu D2 is assigned to the next four positions. Next psu's D3, D4, B1 and B2 follow in sequence and in accordance with their adjusted probabilities. There are no cases of self-pairing of psu's in illustration 5, nor would there be unless unusual conditions prevailed. (One psu might have probability greater than .5.) The ordered assignment of psu's to cell positions continues from stratum I into II and throughout stratum III.

Illustration 3. Equal Sized Strata Constructed from Data in Example Y of Illustration 1 in Preparation for Two Selections per Stratum

Strata	Groups	Classes	Sums of Probabilities for Cells	PSU Codes	PSU Probabilities	Cumulative Sums of Probabilities	Adjusted Probabilities* For PSU's	Adjusted Probabilities* Cumulative Sums
1	2	3	4	5	6	7	8	9
I	1	D	1.5	1	.5	.5	.4	.4
				2	.4	.9	.4	.8
				3	.3	1.2	.3	1.1
				4	.3	1.5	.2	1.3
		B	.7	1	.3	1.8	.3	1.6
				2	.4	2.2	.4	2.0
II	1	A	.2	1	.2	2.4	.2	2.2
	2	A	.5	2	.5	2.9	.6	2.8
		B	.3	1	.3	3.2	.4	3.2
		C	.4	1	.4	3.6	.5	3.7
		D	.2	1	.2	3.8	.3	4.0
III	3	C	.6	1	.2	4.0	.2	4.2
				2	.4	4.4	.4	4.6
		B	.2	1	.2	4.6	.2	4.8
		A	1.4	1	.5	5.1	.4	5.2
				2	.4	5.5	.4	5.6
				3	.3	5.8	.2	5.8
				4	.2	6.0	.2	6.0

* In practice, adjusted probabilities ususally will be within rounding error of initially calculated probabilities.

Illustration 4. A Stratified Random Design with Two Selections per Stratum, Chosen in Proportion to Assigned Probabilities and with Replacement*

Strata	Sample Numbers 1	2	3	4	5	6	7	8	9	10
	Weights .1	.1	.1	.1	.1	.1	.1	.1	.1	.1
I	D1	D2	D2	D1	B2	B2	D3	D2	B1	D4
	D4	B2	D1	D3	B2	B1	D2	B1	D1	D3
II	B1	A2	A2	C1	B1	C1	B1	A2	D1	A2
	A1	A2	A1	C1	C1	C1	D1	D1	A2	B1
III	C2	A2	A2	A4	A4	C1	C1	C2	B1	B1
	C2	A3	A1	A1	A1	C2	A2	A3	A2	A1

* Notice that the number of appearances of each psu agrees exactly with its adjusted probability in Illustration 3. Although such would not be the case with every set of unrestricted random selections, the illustrated formation, a convenient device for variance calculations, was used in calculating the variances shown in Tables 5 and 6.

Illustration 5. An Ordered Systematic Design with Two Selections per Stratum*

Strata	Sample Numbers 1	2	3	4	5	6	7	8	9	10
	Weights .1	.1	.1	.1	.1	.1	.1	.1	.1	.1
I	D1	D1	D1	D1	D2	D2	D2	D2	D3	D3
	D3	D4	D4	B1	B1	B1	B2	B2	B2	B2
II	A1	A1	A2	A2	A2	A2	A2	A2	B1	B1
	B1	B1	C1	C1	C1	C1	C1	D1	D1	D1
III	C1	C1	C2	C2	C2	C2	B1	B1	A1	A1
	A1	A1	A2	A2	A2	A2	A3	A3	A4	A4

* Selection probabilities and the ordering of psu's are those appearing in column 8 of Illustration 3.

REFERENCES

Brooks, Camilla, "The Effect of Controlled Selection on the Between PSU Variance," *Proceedings of the Social Statistics Section*, American Statistical Association, 1975, pp. 336-341.

Goodman, Roe and Kish, Leslie, "Controlled Selection - A Technique in Probability Sampling," *Journal of the American Statistical Association*, Vol. 45, 1950, pp. 350-372.

McNerney, Walter J. et al., *Hospital and Medical Economics*, Hospital and Educational Trust, Chicago, 1962, Vol. 1, pp. 573-576.

U.S. Bureau of the Census, *County and City Data Book, 1962*, U.S. Government Printing Office, Washington, D.C., 1962.

_____, *The Current Population Survey, A Report on Methodology*, Technical Paper No. 7, U.S. Government Printing Office, Washington, D.C., 1963.

_____, *County and City Data Book, 1972*, U.S. Government Printing Office, Washington, D.C., 1973.

_____, *The Current Population Survey, Design and Methodology*, Technical Paper 40, U.S. Government Printing Office, Washington, D.C., 1978.

Wilkerson, Marvin, "The Revised City Sample for the Consumer Price Index," *Monthly Labor Review*, October 1960, Bureau of Labor Statistics, pp. 1078-1083.

ACKNOWLEDGEMENT

The authors acknowledge contributions to this research by Roe Goodman, Judith Connor, Robert Groves, Lou Magilavy, Robert Sinsheimer.

REQUESTS FOR DESCRIPTIONS OF VARIANCE CALCULATION FORMULAS should be addressed to: Sampling Section, Survey Research Center, The University of Michigan, P.O. Box 1248, Ann Arbor, Michigan 48106.