

COST BENEFIT ANALYSIS

CPI 2007 Revision Initiative

Revised 1, June, 2005

Report on Plans for Continuous Revision of CPI Geographic and Housing Samples

Introduction

The Consumer Price Index (CPI) is the principal source of information concerning trends in consumer prices and inflation in the United States, and is one of the Nation's most important economic indicators. The measure is used extensively for economic analysis and policy formulation in both the public and private sectors. The CPI also is used to adjust payments to social security recipients and to federal and military retirees, and for a number of entitlement programs such as food stamps and school lunches. In addition, the CPI is used to adjust individual income tax brackets, exemption amounts, and other tax parameters for changes due to inflation.

In order to maintain the accuracy and currency of the CPI, comprehensive updatings of the CPI have been undertaken by the Bureau of Labor Statistics (BLS) about every ten years. There have been six such Revisions in the history of the CPI. Revision periods provide an opportunity to reflect changes in the geographic distribution of the population and in consumers' buying habits; to incorporate improvements in technology and index methodology; to update survey techniques; and to modernize computer system hardware and software. In the past, Revisions were funded through periodic budget increments. The most recent CPI Revision was funded through a multi-year initiative beginning in FY 1995. In addition to sample updating, it included projects to develop a new housing estimation system, a computer-assisted data collection system, and a new Telephone-based Point of Purchase Survey (TPOPS).

In December 1998, the BLS announced it would update the consumption expenditure weights in the Consumer Price Index for all Urban Consumers (CPI-U) and the CPI for Urban Wage Earners and Clerical Workers (CPI-W) to the 1999-2000 period, effective with release of data for January 2002. Additionally, CPI expenditure weights would be updated at two-year intervals subsequent to the 2002 updating. This policy change represented the first major step in moving from a decennial Revision schedule to a more accelerated process. The 1999-2000 weights, which were introduced in 2002 as scheduled, replaced 1993-95 weights that were first used in the index effective with January 1998 data. The next weight update will occur effective with release of CPI data for January 2004, when the weights will be updated to the 2001-02 period. As a result of this change, expenditure weight data will be, on average, "two years old" when introduced into the CPI, and four years old when replaced. By contrast, the 1993-95 weights were, on average, 3½ years old in January 1998, and they replaced weights that were about 15 years old.

In FY 2002, the BLS received funds to take further steps in revising and updating the Consumer Price Index on a continuous basis. Outlet sample rotation—that is, updating of stores and other establishments in which prices are collected—is now being completed on a four-year rather than the previous five-year cycle. Beginning in FY 2003, item samples for a significant proportion of index categories will be reselected midway between each four-year outlet sample rotation. Continuous modernization also is underway in the computer systems area. Work is now focused on upgrading and improving the infrastructure underlying the commodity and service components of the index.

The last important part of the Continuous Updating program initiated in 2002 was to conduct an evaluation of whether or not the continuous revision process can be extended to revising and updating the samples of geographic areas in which prices are collected and housing units for which rents are collected.¹ This report provides the results obtained thus far from that ongoing evaluation.

The sample of geographic areas on which the index is built is selected on the basis of population data from the Decennial Census. The current CPI geographic sample was based on the 1990 Census and has been used in the index since 1998. In the past, new areas were introduced into the official CPI over a short period of time, necessitating a temporary but sharp increase, or “spike,” in the number of staff and related program resources. For example, during the last revision in the CPI area sample, 36 new areas were introduced in February 1998. In order to carry out the work associated with introducing the new areas, BLS had to increase staff and related resources significantly over a short period of time. Training and managing these resources can be both costly and somewhat wasteful, since the staff levels must be reduced when the revision period concludes. Moreover, the rapid increase in workload can create an environment conducive to data collection or processing errors.

Revising the sample of housing units is the other major Census-based CPI Revision activity that BLS has not yet converted to a continuous process. CPI prices are collected using two surveys: the Housing survey, used for the Residential Rent and Owners’ Equivalent Rent indexes, and the Commodities and Services (C&S) survey used for all other item categories. As discussed above, outlet and item samples for C&S categories are rotated on a regular basis. When the geographic sample changes, the C&S rotation process must be redirected to a different set of areas, with some added complexity arising from the necessity of introducing entire outlet and item samples simultaneously in the new areas. This adds to the overall C&S rotation cost but does not represent a dramatic increase in resource requirements.

In contrast to C&S, there is no current process, and hence no source of funding, for housing sample rotation. Historically, updated samples of housing units for measuring changes in rental values have been introduced about every ten years, along with introduction of the area samples. During the last Revision, for example, housing samples in both the new and continuing CPI areas were introduced into the index in January 1999, one year after the introduction of the area sample. The Decennial Census has provided the necessary information on the location and types of housing units within geographic areas. Elimination of some questions from the 2000 Decennial Census short form have reduced its value in selecting locations within cities in which to collect rent data. A more fundamental deficiency in the use of Census data as a frame, however, is the inability to update the housing sample more frequently than decennially in response to changes in the neighborhood locations and types of housing units in which people live. Although the CPI samples are augmented with samples of newly constructed housing units, this is only a partial solution, not a true sample rotation process.

Housing sample reselection is a very costly activity, and the BLS cannot achieve a level updating budget without identifying a means of rotating housing samples on a continuous basis.

¹ As stated in the Decision Paper for the Continuous Updating initiative, “By FY 2003, the goal is to produce a new plan for carrying out these activities that will include the level of effort and resources required. These planning activities will, of necessity, also address the implications of a continuous CPI revision for the Consumer Expenditure Survey program.”

This is likely to require that alternatives to the Decennial Census be found for sample frame and weighting information, both in new and continuing geographic areas.

It should be emphasized that reselections of the geographic and housing samples are almost inseparable activities from an operational standpoint. In particular, continuous area rotation probably requires that the BLS find a way to select housing samples without reliance on Census data alone. Otherwise, housing samples in the last rotating areas will be initiated long after the Census year on which their housing sample designs are based. This would work against the goal of maintaining timely and representative CPI samples. Meanwhile, it would be impossible to rotate the area sample for C&S pricing while leaving housing pricing in the old area sample. Field workload considerations make it infeasible to maintain collection of housing data in one set of cities and collection of prices of other commodities and services in a different set of cities.

Adding to the task of developing a continuous process for housing sample updating is a dissatisfaction with the operational methods used in the past. It is hoped that alternatives to these methods of locating and initiating rental housing units can be found that are less costly and more effective.

The remainder of this report lays out an approach to continuous rotation area and housing samples in turn, highlighting which planning activities have been accomplished and what research issues have yet to be resolved.

PSU Revision

This section discusses the necessary first step in designing a rotation plan for CPI areas; namely, the selection of the areas that will comprise the new sample². This step is now complete. The new sample will also be employed in the Consumer Expenditure (CE) Survey beginning in 2005. To enable the CE household sample to be selected in coordination with the sampling for other Federal household surveys, it was necessary to provide the list of CPI/CE areas to the Census Bureau in July of 2002.

The new area sample was selected based on the 2000 Census of Population. It was first necessary to determine the basic constraints on the process, in particular the constraints imposed by the CPI area publication structure. Changes in populations between 1990 and 2000 required us to address the publication issue, since the currently published metropolitan areas—selected, as noted above, based on the 1990 Census—are no longer the largest areas in population. It was therefore necessary either to increase the number of published cities, or to drop some that are currently published.

In late FY 2001, joint meetings were initiated among representatives of the CPI and CE programs, as well as the BLS's Division of Price and Index Number Research and Office of Survey Methods Research, to review what was done in the past and what changes might be desired for the future. Discussions focused on whether a reduction in the number of primary sampling units (PSUs), either certainty or non-certainty, could lead to an increase in the accuracy of the CPI. Several scenarios were developed and simulated using detailed cost and variance information taken from the CPI Item-Outlet Optimization Model. Based on these simulations

² Note that this document describes the original 86 area design. The final initiative with a reduced funding level is based on a 75 area design. The costs and benefits for the IT components do not change as a result of this change in area design and the document has not been updated to reflect the smaller area design.

and other considerations, it was decided that the number of published metropolitan areas would be reduced by three, with Milwaukee, Kansas City, and Cincinnati being dropped, while the West D (non-metropolitan) stratum would move from unpublished to published status.

Background. Currently, CPI series are published for 27 metropolitan areas with 1990 populations of at least 1.5 million. (Phoenix was added as the 27th published area in January 2002.) These include the 25 largest areas in 1990, plus Honolulu and Anchorage, which have much smaller populations; publication of the latter two areas is considered justifiable because of their unique locations. In addition, CPI data are published for smaller metropolitan areas (referred to as the B/C strata in CPI publications) in four Census regions and for non-metropolitan (D-size) urban areas in two Census regions.³ Data are collected but not published for the non-metropolitan West region, and the CPI has no D PSUs in the Northeast region because the population in that stratum was so small in 1990.

The total number of CPI PSUs is 87: of these, 31 are in the 27 published metropolitan areas (New York comprises three PSUs and Los Angeles and Washington/Baltimore each comprise two), 46 in the B/C strata and 10 in the D strata. Non-metropolitan, primarily rural PSUs also are sampled in all four regions for the CE program only. Altogether, there are 38 CPI “index areas,” for which basic CPI indexes are computed: the 31 A PSUs, the four regional B/C strata, and the three non-empty regional D strata.

Approximately 87 percent of the 1990 U.S. population were members of households covered by the CPI-U. This share will increase when samples based on the 2000 Census are incorporated in the CPI, both because of changing populations and because of changing OMB area definitions. Notably, the CPI-U population will be defined to include all residents of metropolitan and micropolitan Core Based Statistical Areas (CBSAs). Also, New England CBSAs will be defined using county rather than town boundaries.

The CPI’s D stratum will then be made up of the new “micropolitan” CBSAs. Partly as a consequence of this definitional change, the population of the Northeast D stratum will be large enough to justify selection of PSUs in that stratum. The increased population will not be sufficient to justify publication of the Northeast D stratum, but the West D stratum will be publishable, unlike the current situation.

It should be noted that specification of PSUs for the CPI and CE was based on preliminary OMB area definitions and populations. It was decided that postponing selection until these were declared final would have imposed unacceptable delay in implementing the new geographic sample.

Selection of New PSU Sample. Table 1 shows the 2000 populations of the 27 published metropolitan areas as proportions of the total CPI-U population. The table also shows that two metropolitan areas, Sacramento and San Antonio, now exceed some published areas in population. This made it necessary for the BLS to reassess the list of cities for which separate CPI index series would be calculated and published.

Selection of PSUs was carried out using statistical methods aimed at minimizing index variance. The most critical constraint imposed on this probabilistic selection process was the

³ Prior to the 1998 CPI Revision, non-self-representing metropolitan areas were published in two size strata in each region. These B and C strata were combined in 1998 and designated the B/C strata to facilitate comparison with earlier published data.

specification of published metropolitan areas, which would be selected with certainty. This decision determined the population boundary between the A (certainty area) strata and the non-certainty B/C and D strata. By reducing the number of published A cities, the standard error of the U.S. CPI can be improved, at the cost of reducing the level of geographic index detail provided to the public.

Simulations were used to evaluate several options. The options were distinguished primarily by the set of published certainty cities. One other issue that was investigated in the simulations was whether it would be possible to reduce the number of sampled B/C and D PSUs with only small upward impacts on variance. If this were true, the cost savings from reducing the total number of PSUs potentially could more than offset the variance loss by permitting an increase in the total number of sample quotes. The simulations demonstrated, however, that this hypothesis was false. Eliminating B/C PSUs led to increases in index variance that were far too large to be justifiable on the basis of cost savings.

There was no support for increasing the number of published metropolitan areas beyond the present 27. Besides implying the publication of local area indexes with high variances, such a strategy would increase the overall U.S.-level CPI variance.

It was also decided that ending publication of Anchorage and Honolulu was not appropriate. Those cities' published indexes date from the early 1960s, and their deletion has been rejected in the past. Moreover, examination of price indexes is at least partially supportive of the hypothesis that Anchorage and Honolulu have unique inflation experiences: over the last 10 years their rates of price increase have been among the lowest of all published CPI cities.

The option ultimately selected was to drop three current A PSUs—Milwaukee, Kansas City, and Cincinnati—without adding either Sacramento or San Antonio, the largest metropolitan areas not currently published. The three dropped cities all gained population at a slower rate than the A cities as a whole, and their deletion represents a partial step toward the most efficient possible sample design for the CPI. Simulations showed that this change would slightly reduce the estimated six-month U.S. All Items standard error.

Examination of the population figures in Table 1 shows that there is a particularly large interval between Cincinnati and the next larger A PSU, Portland, providing what can be viewed as a natural point at which to place the A stratum boundary. The 296,000 population difference between Cincinnati and Portland is wider, for example, than the 290,000 range that contains Cincinnati, Sacramento, Kansas City, San Antonio, and Milwaukee. This means that the A boundary could be defined somewhere in this interval, at a population level that would neither narrowly include nor narrowly exclude any city.

This change also reduces the number of index areas, to 36. This offers another potential gain, from the mitigation of any small sample bias that may arise when basic indexes are computed using small numbers of prices. Allocating the same total number of quotes to a smaller number of index areas is likely to reduce the number of small item samples.

Having specified the number of certainty, self-representing PSUs, the CPI program was then able to select the complete set of 86 PSUs for the new, 2000 Census based sample. Initially, the CPI divided the remaining urban population into 58 equivalent strata. Each strata was designated to cover about the as much population as the smallest A-sized city. Based on the population, 42 strata were designated for the medium sized cities (previous B/C-sized cities are

referred to in the table as X-sized) and the remaining 16 were designated for the micropolitan areas (D-sized cities are referred to in the table as Y-sized). Geographic areas were then mapped into the strata based on their population and on their longitude and latitude, variables that had been shown to be most significant in explaining price differential between areas. The result was 58 strata each composed of two or more "price change-equivalent" areas. One PSU was then selected from each strata using a "keyfitted" probability sampling method that increases the probability of reselecting a PSU currently in the sample provided that it has not lost population between the two selection periods. For a more complete description of the sample selection process see the article by William Johnson, Owen Shoemaker, and Yeon Rhee "Redesigning the Consumer Price Index Area Sample," attached as Appendix I.

The new geographic sample will include the 22 largest urban areas, comprising 26 PSUs, plus Anchorage and Honolulu. Another 27 smaller areas from the current sample are included in the new sample as well. Thirty-one areas will be new to the CPI sample, replacing 32 current sample areas.

The areas in the new sample are shown in Table 2. As the table shows, four PSUs will be allocated to the West D stratum, permitting publication of that stratum index for the first time.

Proposed Rotation Schedule. Over the past year the BLS has developed a revised schedule for introducing the new geographic areas into the CPI sample. As noted above, in 1998 all the new areas in the 1990 Census based sample were introduced simultaneously. In the proposed new schedule, only four or five areas will be introduced in any one-year, thereby smoothing out the required resource level. Under this plan, the 31 new areas will be divided into seven groups; the first group will be used in the CPI for the first time in 2008, and the last group will be used for the first time in 2014.

Chart 1 displays the schedule for bringing in the seven PSU groups. In each area, several steps must be completed prior to initial use of price data in the index: TPOPS collection, outlet sample processing, and initiation. As shown in the chart, each new geographic area will roll into the CPI through a four-year process. During the first two years, the BLS will develop a sample of retail establishments and outlets from which to collect prices by working with the Bureau of the Census to conduct a TPOPS survey in the new area. The TPOPS survey will be an extension of our current survey process, which is carried out in every PSU on a continuous four-year rotation cycle. The regular sampling schedule of item and PSU categories will be adjusted to accommodate the need for complete, timely outlet samples in the new areas. In year three, BLS will process the data collected in the TPOPS survey and establish a field presence in the new geographic areas, hiring and training staff. In year four, BLS will initiate pricing of the CPI sample in retail establishments selected to represent the geographic area. At the conclusion of year four, the new geographic area will replace an existing area in the computation of the commodities and services component of the official CPI. At the start of each year, a new set of geographic areas will begin the four-year process. When fully implemented, both Census TPOPS and BLS processing and initiation activities will be underway simultaneously, although in different geographic areas. Tables 3 and 4 below show the activities in more detail and contrast the new process with the existing one. Table 3 addresses adding geographic areas and Table 4 addresses dropping areas.

Table 5 lists the CPI geographic areas in order of their priority for introduction (deletion) to (from) the CPI. PSU's were grouped into four prioritized categories based on the importance

of getting the new geographic areas into the CPI. Within the new geographic areas, the first category, referred to as "geographic holes," are the most important to get into the index. These nine new geographic areas are those selected to represent strata for which there is no priced area in the current sample design. For example, Augusta, Maine and Ithaca, New York are small cities in the Northeast, a stratum of the population not currently represented in the CPI geographic sample. It is important to add these areas as quickly as possible in order to represent fully the U.S. urban population. The ten geographic areas in category 1 that can be dropped immediately are labeled "strata duplicates." They can be dropped because they were not reselected to continue in the CPI and are in geographic strata for which another continuing area is present. For example, Johnstown, Pennsylvania is in the same strata as Reading, Pennsylvania. As noted above, Johnstown was not reselected from the 2000 Decennial CPI area update, while Reading was. Because they are both in the same geographic stratum, it is unnecessary, and in fact inefficient, to price both of these areas.

Each of the remaining three categories is comprised of sets of matched new and dropping geographic areas. The transition from the old sample to the new for these areas will be accomplished through the matched adding and dropping of areas simultaneously. Each category is prioritized based on the quality of the pairings. For example, the second group ("150-mile similar match") is comprised of five new geographic areas matched with five dropping areas that are within 150 miles of one another. These are the least similar matches and therefore are the highest priority in adding and dropping. For example, we can continue to price in Chanute, Kansas until we are able to add Springfield, Missouri, although Chanute is a small or C-sized area while Springfield is a medium sized or B-sized one. Next in priority are the three index area matches. For each of these three new geographic areas, there currently exists an area that can remain in the CPI as a proxy until the new area is introduced. While the proxy area is not from the same strata as the new, the match is similar in terms of index, or publication, area. For example, Gainesville, Florida is in the same publication area, medium-sized South, as Jacksonville, Florida. Finally, there are 14 pairs of strata matches. In these cases, it is relatively unimportant when they are rotated because the new area and the dropping area are both from the same strata, and therefore which area to price is a stochastic event. For example, it is a matter of chance whether we selected Dayton or Bellefontaine, Ohio for pricing as both are part of the Y-208 strata in the North Central U.S.

In addition to adding new areas, converting to the new sample will include redefining the geographic boundaries of continuing PSUs to conform to the updated OMB definitions, as counties are added to PSUs or moved from one PSU to another. For example, Hampden and Hampshire counties in Massachusetts were moved from the Springfield, Massachusetts PSU to the Boston SMSA, leaving Franklin County as the residual component of Springfield. The CPI introduces these geographic changes as quickly as possible, as part of the sample rotation. For expenditure weights, based on data from the Consumer Expenditure Survey, the new geographic definitions are fully reflected in the CPI beginning with the data for January 2008. For TPOPS, the conversion to the new geographic definitions will take place such that all samples drawn for initiation and use in the CPI after January 2008 will reflect the new geography. It should be noted that until the remaining C&S samples are rotated out of the CPI they will be based on the old geographic definitions. Housing samples will similarly be updated so that new samples initiated in anticipation of the January 2008 index will reflect the new definitions, while older samples will reflect current definitions until updated through replacement or rotation.

Tradeoff of Timeliness and Efficiency. The new approach allows staff to be retained longer, leading to a more efficient and skilled workforce. The corresponding disadvantage is that it will take longer than in the past to complete the area sample rotation. To the extent that rapid, significant population shifts have taken place or will take place over the next several years, the efficiency of the CPI could be reduced by the longer retention of areas that were selected on the basis of the 1990 Census.

It should be emphasized, however, that every two years the BLS updates the expenditure weights attached to item-area categories in the CPI. As part of those biennial updates, the population weights for CPI sample index areas are updated. Thus, the loss in efficiency from retaining PSUs longer than would otherwise be the case comes only from the fact that the retained PSUs may provide less efficient estimates of spending patterns for the index areas that they represent. For example, until rotated in 2011, the index for Stratum X364 will be estimated based on expenditure patterns and prices that are reflective of Gainesville (stratum X350). While both areas represent southern X-sized cities, their division into separate strata indicates that they do not share the same price-change-determining characteristics. Another unavoidable feature of the proposed PSU rotation schedule is that there will be a temporary increase in sampling error during 2008, resulting from the existence of nine PSU strata that contain no current CPI PSUs. Four of the nine new PSUs in those strata cannot be introduced efficiently until 2009. It will be prudent, therefore, to provide for a slight increase in item and outlet sample sizes to ensure that the resource-efficiency benefits of the new plan are not offset by any deterioration in estimation accuracy. Design simulations based on not having two PSUs in X300 and two PSUs in X400 revealed that the CPI could expect an increase in 6-month percent change standard error of about 1.95 percent, from 0.103 to 0.105, and a variance increase of 3.94 percent, from .01065 to .01107. This loss of efficiency due to not having the four “holes” in would require about a five percent increase in overall sample size.

Future PSU Sample Updates. After 2014, it is expected that another sample will begin to be introduced into the CPI on a rolling or continuous basis. It is likely that, following past practice, the 2010 Census will provide the population data underlying this next rotation. If an alternative source of data becomes available on a continuous basis, however, that source may become the basis for the post-2014 rotations. The Interagency sample redesign task force (the Consortium) has recently begun an examination of sample selection methodologies that would place less reliance on the decennial but would likely require full funding for the American Community Survey.

Relationship to CE Sample. The CE survey will introduce the new geographic sample in 2005. This means that the 2005-2006 expenditure weights used in the CPI-U and CPI-W beginning in 2008 will be based on the new sample. Final index values of the Chained CPI for All Urban Consumers, or C-CPI-U, are based on current expenditure data, so the CPI will need to develop a reverse mapping so that the current geographic sample will be reflected in the weights used in the final C-CPI-U indexes for 2005 through 2008. We are currently discussing whether, and how, the CE could adopt a rolling geographic rotation of its household samples in the future. A final decision on this will depend in part on the work currently underway through the Consortium.

Housing Revision

Background. The CPI housing sample is a sample of rental housing units that supports the two largest item categories in the index, Residential Rent and Owners' Equivalent Rent. Revision of the housing sample, like geographic sample revision, historically has taken place at approximately ten-year intervals using Decennial Census data. The current housing samples were selected based on 1990 Census data. The samples were introduced with the January 1999 CPI, simultaneously with introduction of a new housing index estimation system and a new computer-assisted data collection methodology (CADC). Since these new samples were introduced in 1999, they have been augmented with "new construction" samples drawn using data on post-1990 construction permit data obtained from the Bureau of the Census. This process attempts to keep the sample representative over time.

Revising the CPI housing sample involves selecting a sample of rental units for each new CPI geographic area (PSU) as well as replacing rental samples in each continuing area. The process that was used to select rental units for the current CPI sample is described in the article by Frank Ptacek and Robert Baskin, "Revision of the CPI housing sample and estimators," attached as Appendix II. First, each of the CPI geographic areas was divided into small pieces of geography, called segments. Data from the Decennial Census short form provided segment-level information from which expenditure estimates were developed for renters and owners. This information was then used to select statistically representative samples of segments.

The next step in the sampling process was to select a random sample of renters in each segment. The Census information on individual occupant households is not available to BLS, however, due to confidentiality rules. As a result, CPI staff had to develop a method for selecting rental unit samples without reference to the Census micro data on housing tenure.

BLS began by having field staff compile a listing of all the addresses in each selected segment by canvassing the neighborhood and physically recording each address. From each listing, a subset of housing units was selected for a personal visit process called screening, which determined whether the units were occupied by renters or owners. Units occupied by renters and otherwise eligible for inclusion in the sample were then initiated for ongoing collection of rent.

The process of listing and screening, which was conducted in 1997 and 1998, was very costly and in many cases it did not produce the desired result. In particular, in areas dominated by owners the number of renters found fell far below expectations. The specific causes for this low yield are not completely clear, although timing likely played an important role. The process began in 1997, long after the end of the 1990 reference period. The delay was unavoidable and stemmed from a number of factors—Decennial collection and processing, selection of PSUs, and delivery of Census data for segment creation and selection. During this time, however, household movements and overall shifts in tenure toward homeownership may have changed the 1990 data on which the sample was based. Other possible contributing factors are inaccuracies in Census data at the low level used by BLS and errors by BLS staff in the use of Census data.

Purchased Housing Lists. Because of these past problems, BLS decided to evaluate new approach for selecting renters for the CPI. The major feature of the new approach is the use of address lists and tenure information that have been developed—and that are updated regularly—in the private sector. The idea is to replace in whole or in part the listing and screening process used by BLS in the past. BLS conducted a study to identify providers of address lists and the

characteristics of those lists. The first issue of concern is whether the coverage of the lists is adequate. A determination needs to be made on whether or not the lists have enough coverage at the geographic level required by BLS so that taking a sample from them would be representative of the universe of housing units.

BLS identified three vendors whose lists BLS wanted to evaluate further. The address lists for these vendors also contained qualitative information concerning housing tenure. The BLS contracted with Westat, a private statistical consulting firm, to evaluate the lists for two geographic areas, Richmond and Baltimore. The evaluation consisted of two parts. The first concerned the accuracy of the address lists in terms of the numbers of addresses relative to the Census. The second part was to conduct telephone and personal interviews to assess the accuracy of individual addresses on the lists, as well as to identify tenure and compare the results to the lists. One of the three vendors was eliminated from consideration prior to the evaluation since Westat determined that their list data did not contain up-to-date geographic information.

The Westat analysis with respect to coverage is attached as Appendix III. Westat compared the numbers of addresses from each of two vendors at the Census block level and compared those block-level counts to the 2000 Census file. They concluded that the gross coverage rate was relatively good for both vendors, but was better for the lists from one vendor, Marketing Systems Group (MSG). At the block level, the MSG information on tenure also appeared to be more accurate.

Appendix IV presents the second part of the Westat evaluation, concerned the accuracy of the reported tenure status. The MSG housing tenure information is in the form of codes. Each address has a code ranging from 0 to 9. A 0 means there is a high probability that the unit is occupied by a renter. A 9 means there is a high probability the unit is occupied by an owner. Westat focused on units in low-renter areas and confined the analysis to units with codes of 0-8, because of the CPI's over-riding need to use the lists to identify renters and because our previous work had verified that values of 9 reliably identified owners. Again, the MSG lists appeared to out-perform the lists from the other vendor. About 8 percent of the addresses on the lists could not be located and some percentage of the units had incorrect tenure status. The quality of the list information was sufficiently high to warrant further consideration, however.

Based on the results of the Richmond and Baltimore evaluations, BLS decided to broaden this line of analysis. Westat will be asked to obtain and analyze address list data for about 550 segments from a representative sample of that portion of the 1990 BLS sample that is continuing. For Richmond and Baltimore we focused on areas with a renter percentage of 40 percent or less. This time the sample of segments will be representative of all types of segments.

The vendors for this analysis will be slightly different from the first study. Only MSG lists will be included; the other vendor will be dropped based on the findings by Westat mentioned above. In addition, address lists from another company, ADVO, will be included in the analysis. ADVO produces lists that contain addresses only. Their lists are purported to be the most complete lists of deliverable addresses available commercially.

The plan is to compare the two lists at the block group level both in terms of the numbers of addresses and the degree of matching or compatibility between them. The address lists produced from our listing operation will also be compared with the two private sector lists. We will be trying to find out if it is necessary to start with ADVO list in order to have a complete frame of addresses for sampling.

The other major goal of this study will be to compare the tenure information from MSG with data from the screening part of our housing survey. BLS is also in the process of obtaining summary level data for twenty-nine CPI pricing areas. The summary data will be counts of addresses for every zip code in each of the 29 PSUs. The listing will contain data for the MSG, ADVO and 2000 Census samples updated by a company named Claritas. The data will be used to evaluate the adequacy of the coverage of the lists over a large portion of the CPI area sample. **Other Housing Issues.** In addition to the use of private-sector data as an alternative to the Census for rental unit sampling, the CPI program is evaluating several other issues in designing the revised housing sample. One concerns the sampling and weighting of segments for the survey. In the last Revision, probability-proportional-to-size (pps) sampling was employed where $s = \text{expenditure}$. However, the 2000 Decennial Census does not provide block-level estimates of expenditure, nor will the Census Bureau's American Community Survey (ACS). The BLS is considering ways to produce block-level expenditures from block-group or tract-level data and is also considering sampling at the block group level rather than the block level in the current sample design. Before that can be done, however, we must derive comparable rent and rental equivalence measures from Census rents and home values.

In the last Revision, home values were regressed against monthly rent values to obtain monthly rental equivalence values. This process appears to be flawed in two ways. First the model chosen imposed a "cap" on the rental equivalence values produced. Second, rented units likely have different values from owned units in the same segment. The BLS is currently exploring alternative ways of formulating monthly rental equivalence values: for example, a user cost approach. Another alternative is to fall back on pps sampling with $\text{size} = \text{number of units}$, as was done in the 1987 CPI Revision process. However, expenditures are the appropriate measure of size for the CPI, and the average level of expenditures per unit is different for owners and renters. According to the Census, for example, the ratio of renters to owners is roughly 1 to 2. According to weights derived from the CE Survey, the ratio of rental expenditure to owner expenditure is roughly 1 to 3. If we decide to use the 1987 approach we will need to modify the process to account for this difference.

Research also is ongoing to determine the housing unit attributes that most directly predict rent change. This could be used to justify the expanded use of "helper segments," which are employed in the current housing sample to represent segments with extremely low numbers of renters to sample, or to provide guidance in selecting a stratification scheme for the new sample design.

Housing Rotation. A major goal of the CPI program is to update the rental unit samples within metropolitan areas on a faster than decennial basis. Most CPI outlet samples are rotated every four years, and starting in 2003 many item samples will be reselected midway between outlet sample rotations—that is, every two years. Housing rotation has remained on a once-a-decade frequency because of the absence of more timely sampling frames and data to use for sample weighting. Lists may provide an alternative sampling frame. Meanwhile, the ACS, if funded by Congress, could provide local area rent and housing value data on a continuous basis.

The BLS currently envisions a six-year rotation cycle for housing samples within PSUs, once the introduction of the new and continuing PSU housing samples is completed. This rotation cycle would correspond to the six collection panels that are now used in the CPI rental unit sample: each panel is now priced two times per year, and in the future we expect that we would reselect one of these panels each year in each PSU.

We are currently looking at the ACS as the source of updating our weights on an ongoing basis. If funding is obtained for FY 2004 and beyond, the ACS will provide estimates suitable to our needs in 2008.

If we use the ACS (or, for that matter, any other source of weighting data), a method of updating weights of the existing sample, particularly in non-self representing PSUs, will need to be devised. During 2003 the CPI program will test alternative sources and approaches for updating the weights. While the solution is not readily apparent at this time, we feel confident that an approach can be developed that will facilitate an ongoing rotation of the housing sample.

Possible Housing Rotation Plan. Chart 2 shows one potential plan for selecting housing samples in new and continuing index areas, and for subsequently rotating those samples on a six-year cycle.

As depicted above, the plan is very similar to the geographic plan presented in Chart 1. For the new geographic areas, we will begin with a complete new housing sample. Similar to the situation with C&S samples, it will several years to bring the new sample into the CPI in each area. During the first year of the new area housing process, we will procure the lists for the segments selected for the CPI for those areas. During the second year, addresses will be selected and field economists will screen the selected units and begin the initiation stage. After initiation, the units will be priced semi-annually for use in the CPI. As is the case with the C&S sample, a new group of geographic areas will begin the process each year. Also as is the C&S sample, some activity will be underway in each PSU.

For the continuing areas, a rotation process will be undertaken that again looks similar to the current C&S process, where one-fourth of the sample is updated in each PSU annually. For housing however, one panel (one-sixth) of the sample will be rotated each year. The rotation process will include the same set of steps as the process for establishing a sample for new geographic areas: identifying strata; procuring lists of housing units; screening for renters; initiating the new rent sample; and pricing and use in the CPI.

Conclusion

In the past year, considerable progress has been made on developing a plan for introducing the new sample of geographic areas and revising the housing sample. A new approach has been developed for adding new PSUs and dropping PSUs no longer needed. The approach spreads the workload over a period of about ten years and eliminates the resource spikes characteristic of previous revisions. At the same time those new PSUs that are needed to reflect the 2000 population distribution will be introduced first. Other PSUs that are merely replacing PSUs in the current sample will be introduced in the out years. Work has begun on the cost of the new approach, but final estimates will depend on the operational aspects of the new housing sample design.

Work on the new housing sample is also progressing. It now appears likely that it will be possible to make important improvements in the design of the new housing sample. In the first half of 2003 we expect to complete work that will result in a decision that address lists available in the private sector can be used both to list housing units and to streamline the screening process for finding renters from which to collect rents on an ongoing basis. It is anticipated that such a

result will free up resources that can be used to improve the accuracy of the CPI. Equally important, it will provide a source of data that can be used to update the sample in each PSU on a regular basis. The other data needed to establish a regular updating process are data on expenditures for rent and rental equivalence. We plan to study ways in which the American Community Survey can be used to generate such data.

The current schedule calls for completing design work for the new housing sample by September 1, 2003. Detailed cost estimates for introducing the revised PSU sample and the new housing sample, including updating, will be completed by the end of the 2003 in time for the 2006 budget cycle. This schedule will provide enough time to permit introduction of new geographic areas in 2008, the date when CE expenditure data for the new PSU sample will be first used in compilation of the official CPI.

Table 1. Population of Largest Metropolitan Area PSUs
(2000 CBSA Definitions)

PSU	CPI Code	Population in 2000	Percent of CBSA Population
Los Angeles	A419	12,365,627	4.8%
Chicago	A207	9,172,106	3.6%
New York	A109	8,008,278	3.1%
New York suburbs	A110	7,718,773	3.0%
Boston	A103	7,098,363	2.8%
San Francisco	A422	7,039,362	2.7%
New Jersey suburbs	A111	6,708,052	2.6%
Philadelphia	A102	6,188,463	2.4%
Detroit	A208	5,456,428	2.1%
Dallas	A316	5,275,921	2.1%
Washington DC	A312	5,027,797	2.0%
Houston	A318	4,715,407	1.8%
Atlanta	A319	4,201,220	1.6%
Los Angeles suburbs	A420	4,008,018	1.6%
Miami	A320	3,876,380	1.5%
Seattle	A423	3,554,760	1.4%
Phoenix	A429	3,251,876	1.3%
Minneapolis	A211	3,136,198	1.2%
Cleveland	A210	2,945,831	1.1%
San Diego	A424	2,813,833	1.1%
St. Louis	A209	2,693,603	1.0%
Denver	A433	2,629,980	1.0%
Baltimore	A313	2,552,994	1.0%
Pittsburgh	A104	2,431,087	0.9%
Tampa	A321	2,395,997	0.9%
Portland	A425	2,275,095	0.9%
Cincinnati	A213	1,979,202	0.8%
Sacramento (New)		1,838,116	0.7%
Kansas City	A214	1,776,062	0.7%
San Antonio (New)		1,711,703	0.7%
Milwaukee	A212	1,689,572	0.7%
Honolulu	A426	876,156	0.3%
Anchorage	A427	319,605	0.1%
Total of Above		137,731,865	53.6%
All Other CBSAs		119,278,302	46.4%
Total CPI-U Population		257,010,167	100.0%

Table 2. New PSUs by Index Area

Publication Area	PSU Name	2008 INDEX AREA CODE	2000 Decennial Population
1	Philadelphia-Wilmington-Atlantic City, PA-DE-NJ	A102	6,188,463
2	Boston-Brockton-Nashua, MA-NH-ME-CT	A103	7,098,363
3	Pittsburgh, PA	A104	2,431,087
4	New York, NY	A109	8,008,278
5	New York-Connecticut suburbs	A110	7,718,773
6	New Jersey-Pennsylvania suburbs	A111	6,661,750
7	Chicago-Gary-Kenosha, IL-IN-WI	A207	9,172,106
8	Detroit-Ann Arbor-Flint, MI	A208	5,456,428
9	St. Louis, MO-IL	A209	2,693,603
10	Cleveland-Akron, OH	A210	2,945,831
11	Minneapolis-St.Paul, MN-WI	A211	3,136,198
12	Washington, DC-MD-VA-WV	A312	5,027,797
13	Baltimore, MD	A313	2,552,994
14	Dallas-Fort Worth, TX	A316	5,275,921
15	Houston-Galveston-Brazoria, TX	A318	4,715,407
16	Atlanta, GA	A319	4,201,220
17	Miami-Fort Lauderdale, FL	A320	3,876,380
18	Tampa-St. Petersburg-Clearwater, FL	A321	2,395,997
19	Los Angeles County, CA	A419	12,365,627
20	Los Angeles suburbs, CA	A420	4,008,018
21	San Francisco, CA	A422	7,039,362
22	Seattle-Tacoma-Bremerton, WA	A423	3,554,760
23	San Diego, CA	A424	2,813,833
24	Portland-Salem, OR-WA	A425	2,275,095
25	Honolulu, HI	A426	876,156
26	Anchorage, AK	A427	319,605
27	Phoenix-Mesa, AZ	A429	3,251,876
28	Denver-Boulder-Greeley, CO	A433	2,629,980
29	Northeast X's		10,891,754
	Providence, RI	X100	
	Reading, PA	X100	
	Syracuse, NY	X100	
	Sharon, PA	X100	
30	North Central X's		24,774,378
	South Bend, IN	X200	
	Rochester, MN	X200	
	Springfield, MO	X200	

	Madison, WI	X200	
	Milwaukee-Racine, WI	X200	
	Cincinnati-Hamilton, OH-KY-IN	X200	
	Decatur, IL	X200	
	Lincoln, NE	X200	
	Elkhart-Goshen, IN	X200	
	Kansas City, MO-KS	X200	
	Saginaw-BayCity-Midland, MI	X200	
	Youngstown-Warren, OH	X200	
31	South X's		47,517,342
	Tulsa, OK	X300	
	Roanoke, VA	X300	
	Louisville, KY	X300	
	Clarksville, TN	X300	
	New Orleans, LA	X300	
	Knoxville, TN	X300	
	Tuscaloosa, AL	X300	
	Fort Hood, TX	X300	
	Jacksonville, FL	X300	
	El Paso, TX	X300	
	SanAntonio, TX	X300	
	BatonRouge, LA	X300	
	Greenville-Spartanburg-Anderson, SC	X300	
	Norfolk-Virginia Beach-Newport News, VA-NC	X300	
	Ocala, FL	X300	
	FortMyers-CapeCoral, FL	X300	
	Florence, SC	X300	
	Birmingham, AL	X300	
32	West X's		15,944,435
	Sacramento, CA	X499	
	BoiseCity, ID	X499	
	LasVegas, NV-AZ	X499	
	Bellingham, WA	X499	
	Fresno, CA	X499	
	Merced, CA	X499	
	Provo-Orem, UT	X499	
	Yuma, AZ	X499	
33	Northeast Y's		2,942,759
	Augusta, ME	Y100	
	Ithaca, NY	Y100	
34	North Central Y's		8,717,815

	Whitewater, WI	Y200	
	Bellefontaine, OH	Y200	
	Brookings-Madison, SD	Y200	
	Macomb, IL	Y200	
35	South Y's		12,322,746
	Valdosta, GA	Y300	
	Henderson, NC	Y300	
	Eagle Pass, TX	Y300	
	Picayune, MS	Y300	
	Winchester, VA	Y300	
	Greenwood, MS	Y300	
36	West Y's		5,274,554
	Newport, OR	Y400	
	Bend-Redmond, OR	Y400	
	El Centro, CA	Y400	
	Prescott, AZ	Y400	

Note: PSUs in bold text are new selections; non-bolded PSUs are continuing.

Table 3. New Geographic Areas

Process	New Process	Current Process
TPOPS at CENSUS	<i>Years 1 and 2</i> Quarterly Survey conducted by Census; 12.5% of all POPS categories sampled in a given quarter. 2 years (8 quarters) required to complete a full TPOPS sample.	Quarterly Survey conducted by Census; 6.25% of all POPS categories sampled in a given quarter. 4 years (16 quarters) required to complete a full TPOPS sample
TPOPS Processing at BLS	<i>Years 2 and 3</i> Quarterly data will be stockpiled until the full TPOPS sample is received. All other processing is identical to current process	Data are stockpiled until 2 quarters are available. Processing includes outlier review and address coding and collapsing and outlet/Item sample selection
Pre-Initiation processing at BLS (Field)	<i>Year 3</i> Outlet/Item Samples will be processed for the entire geographic area at once. Otherwise the process is identical to the current process.	Field activities include parsing of samples into individual EA assignments, collapsing to existing outlets and refining address and contact information.
Initiation of Sample in Field	<i>Year 4</i> The entire outlet/Item sample will be initiated at the same time	12.5% of the geographic area's outlet/item sample is rotated each half-year. (Item-Outlet Rotation) In addition, 12.5% of area's item samples are updated in the existing outlets. (Within outlet rotation)
Pricing	<i>Year 5</i> Same as current process.	The entire Outlet/Item sample is priced according to the pricing schedule assignment (either monthly or bimonthly)

Table 4. Dropping Geographic Areas

Process	New Process	Current Process
TPOPS at CENSUS	<i>Year 1</i> All TPOPS data collection will cease at Census 48 months (16 quarters) prior to the date at which a geographic area will drop from the official CPI	Quarterly Survey conducted by Census; 6.25% of all POPS categories sampled in a given quarter. 4-years required to complete a full TPOPS sample
TPOPS Processing at BLS	<i>Years 1 and 2</i> Data collected prior to the cessation of data collection at Census will be processed per the existing procedures	Data are stockpiled until 2 quarters are available. Processing includes outlier review and address coding and collapsing and outlet/Item sample selection
Pre-Initiation processing at BLS (Field)	<i>Years 2 and 3</i> Data collected prior to the cessation of data collection at Census will be processed per the existing procedures	Field activities include parsing of samples into individual EA assignments, collapsing to existing outlets and refining address and contact information.
Initiation of Sample in Field	<i>Years 2 and 3</i> Outlet/Item samples selected from TPOPS Data collected prior to the cessation of data collection at Census will be initiated in the field per the existing procedures	12% of the geographic area's outlet/item sample is rotated each half-year
Pricing	<i>Year 5</i> All price data collection will cease in January of the drop year.	The entire Outlet/Item sample is priced according to the pricing schedule assignment (either monthly or bimonthly)

Table 5. New Areas by Priority of Introduction

	Added Areas		Dropping Areas
Group 1	Geographic holes		Strata Duplicates
	Y102	Augusta ME	Springfield, MA
	Y104	Ithaca NY	Buffalo, NY
	Y430	El Centro CA	Burlington, VT
	Y432	Prescott AZ	Johnstown, PA
	Y206	Whitewater WI	Albany, GA
	X342	Louisville KY	Brownsville, TX
	X354	New Orleans LA	Amarillo, TX
	X476	Bellingham WA	Melbourne, FL
	X480	Merced CA	Faribault, MN
			Statesboro, GA
Group 2	150-mile "similar" Match		
	X348	Clarksville TN-KY	Evansville, IN
	X230	Springfield, MO	Chanute, KS
	Y316	Henderson NC	Richmond, VA
	Y324	Greenwood MS	Pine Bluff, AR
	X106	Providence RI	Hartford
Group 3	Index-Area Matches		
	X364	Jacksonville FL	Gainesville, FL
	X362	Fort Hood TX	Beaumont, TX
	X478	Fresno CA	Modesto, CA
Group 4	Strata Matches		
	Y208	Bellefontaine OH	Dayton, OH
	Y322	Winchester VA	Morristown, TN
	Y318	Eagle Pass TX	Lafayette, LA
	X336	Roanoke VA	Raleigh, NC
	X360	Tuscaloosa AL	Florence, AL
	X214	South Bend IN	Columbus, OH
	Y314	Valdosta GA	Arcadia, FL
	Y212	Macomb IL	Mt. Vernon, IL
	X334	Tulsa OK	Oklahoma City, OK
	X470	Sacramento CA	Chico, CA
	X358	Knoxville TN	Chattanooga, TN
	X216	Rochester MN	Wausau, WI
	X366	El Paso TX	Midland, TX
	Y426	Newport OR	Pullman, WA

Chart 1. PSU Rotation PLAN

Group	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Group 1 Adding 5 PSUs	Census TPOP collection		processing	INITIATION	Pricing						
Group 1 Dropping 10 PSUs	No TPOPS collection		no processing or initiation		Drop PSUs						
Group 2 Adding 4 PSUs	Census TPOP collection		processing	INITIATION	Pricing						
Group 3 Adding 5 PSUs		Census TPOP collection		processing	INITIATION	Pricing					
Group 3 Dropping 5 PSUs		No TPOPS collection		no processing or initiation		Drop PSUs					
Group 4 Adding 5 PSUs			Census TPOP collection		processing	INITIATION	Pricing				
Group 4 Dropping 5 PSUs			No TPOPS collection		no processing or initiation		Drop PSUs				
Group 5 Adding 4 PSUs				Census TPOP collection		processing	INITIATION	Pricing			
Group 5 Dropping 4 PSUs				No TPOPS collection		no processing or initiation		Drop PSUs			
Group 6 Adding 4 PSUs					Census TPOP collection		processing	INITIATION	Pricing		
Group 6 Dropping 4 PSUs					No TPOPS collection		no processing or initiation		Drop PSUs		
Group 7 Adding 4 PSUs						Census TPOP collection		processing	INITIATION	Pricing	
Group 7 Dropping 4 PSUs						No TPOPS collection		no processing or initiation		Drop PSUs	

Chart 2. PSU Housing Rotation PLAN											
Group	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Group 1 Adding 5 PSU's		Procure List	Screen	INITIATION	Pricing	Segments	Procure List	Screen	INITIATION	Pricing	Segments
Group 1 Dropping 10 PSU's					Drop PSU's						
Group 2 Adding 4 PSU's			Procure List	Screen	INITIATION	Pricing	Segments	Procure List	Screen	INITIATION	Pricing
					Drop PSU's						
Group 3 Adding 5 PSU's				Procure List	Screen	INITIATION	Pricing	Segments	Procure List	Screen	INITIATION
Group 3 Dropping 5 PSU's							Drop PSU's				
Group 4 Adding 5 PSU's					Procure List	Screen	INITIATION	Pricing	Segments	Procure List	Screen
Group 4 Dropping 5 PSU's							Drop PSU's				
Group 5 Adding 4 PSU's						Procure List	Screen	INITIATION	Pricing	Segments	Procure List
Group 5 Dropping 4 PSU's								Drop PSU's			
Group 6 Adding 4 PSU's							Procure List	Screen	INITIATION	Pricing	Segments
Group 6 Dropping 4 PSU's									Drop PSU's		
Group 7 Adding 4 PSU's								Procure List	Screen	INITIATION	Pricing
Group 7 Dropping 4 PSU's										Drop PSU's	
55 Ongoing PSU's Panel 1		Segments	Procure List	Screen	INITIATION	Pricing		Segments	Procure List	Screen	INITIATION
55 Ongoing PSU's Panel 2			Segments	Procure List	Screen	INITIATION	Pricing		Segments	Procure List	Screen
55 Ongoing PSU's Panel 3				Segments	Procure List	Screen	INITIATION	Pricing		Segments	Procure List
55 Ongoing PSU's Panel 4					Segments	Procure List	Screen	INITIATION	Pricing		Segments
55 Ongoing PSU's Panel 5						Segments	Procure List	Screen	INITIATION	Pricing	Pricing
55 Ongoing PSU's Panel 6							Segments	Procure List	Screen	INITIATION	Pricing

List of Appendices

- I. Johnson, William H., Owen J. Shoemaker, and Yeon W. Rhee (2002), "Redesigning the Consumer Price Index Area Sample," *Proceedings of the Section on Government Statistics*, American Statistical Association.
- II. Ptacek, Frank, and Robert M. Baskin, (1996), "Revision of the CPI housing sample and estimators," *Monthly Labor Review*, December.
- III. Westat, (September 25, 2002), *Evaluation of Vendor Lists Using Census Data*.
- IV. Westat, (February 26, 2003), *Evaluation of Vendor Lists: Comparison of Vendor and Interview Data*.

Redesigning the Consumer Price Index Area Sample

William H. Johnson, Owen J. Shoemaker, and Yeon W. Rhee
U. S. Bureau of Labor Statistics, 2 Mass Ave NE, Room 3655, Washington, DC 20212

KEY WORDS: multistage, stratified, controlled selection, overlap

Any opinions expressed in this paper are those of the authors and do not constitute policy of the Bureau of Labor Statistics.

This paper describes the PSU selection process for the next CPI Revision. The U. S. Consumer Price Index (CPI) employs a multistage sample design that has been revised every ten years. The first stage consists of selecting primary sampling units (PSUs) which are formed from Metropolitan or Micropolitan Core Based Statistical Areas (CBSAs) based on preliminary definitions by the Office of Management and Budget.

The PSU selection process for the next CPI Revision is quite similar to the process of selecting the sample for the 1998 CPI Revision (see Williams et al). The biggest difference has been the use of variance models of six-month index change for the Commodities and Services part of the CPI-U in determining the set of certainty PSUs and the distribution of non-certainty PSUs across Census region by size class combinations. Alternative methodologies for stratifying PSUs prior to selection were considered and work on modeling CPI-U change since 1992 influenced the selection of stratifying variables. All of the programs involved in the work on selecting the 1998 CPI Revision PSU sample were updated or rewritten.

The process of selecting the PSU sample involves six steps:

- 1) Determine the PSUs selected with certainty
- 2) Determine the number of non-certainty PSUs and their distribution across regions
- 3) Stratify the non-certainty PSUs
- 4) Use Keyfitzing to improve expected overlap

- 5) Use controlled selection to generate a set of sampling patterns and weights
- 6) Select a sample of PSUs

Determining Certainty PSUs

The first step in the process of selecting the PSU sample is to determine which PSUs are certainty PSUs. In order to determine the certainty PSUs it was necessary to determine the possible certainty PSUs. The most likely certainty PSUs are those which are already certainty PSUs in the existing CPI area sample. However with the shift to CBSA based definitions it became necessary to determine what the new definitions of the current certainty PSUs are likely to be. The certainty cities were mapped along with preliminary CBSA definitions. It was assumed that a CBSA would either be entirely included or entirely excluded from these areas. In cases where a CBSA was partially contained in a current certainty PSU, the probability of the outside counties being in the final definition given to BLS by the Census Bureau was examined as part of the assessment of whether to include or exclude the CBSA.

After the expected definitions of the current certainty cities were decided, the remaining possible certainty cities were the remaining individual metropolitan CBSAs. The largest metropolitan CBSAs outside of the current certainty cities were determined and considered for inclusion in the list of new certainty PSUs.

Next it was necessary to determine the criteria for PSUs to be selected with certainty. There were several possible options. The entire CPI-U population to be represented is the total population contained in all metropolitan and

micropolitan CBSAs. This population is 257,010,167.

The options considered included:

- 1) 1,500,000 – the population cutoff used previously for determining certainty cities
- 2) 1,680,000 – a population cutoff that wouldn't cause the loss of any current certainty cities
- 3) 1,800,000 – a population cutoff which was considered for use previously
- 4) 2,141,751 – the population cutoff obtained by using 120 half sample equivalents (HSEs) to represent the total population of 257,010,167
- 5) 2,570,102 – the population cutoff obtained by using 100 HSEs to represent the population of 257,010,167
- 6) 4,283,501 – the population cutoff obtained by using 60 HSEs to represent the population of 257,010,167

A half sample equivalent is a unit of sample size. Each certainty city will receive at least two HSEs and each selected non-certainty city will receive one HSE.

The option of using 1,500,000 as a population cutoff for determining certainty cities was dropped as it would add too many certainty cities to be affordable. Each certainty city must have enough sample for their individual city CPIs to be publishable on at least a semi-annual basis. This makes the certainty cities much more costly than non-certainty cities.

The decision as to which set of cities should be selected with certainty required information so one could compare the various possible sets of certainty PSUs. In order to compare the various options, the model used for optimizing the CPI Commodities and Services sample was generalized. (see Leaver et al) This model attempts to select outlet and item sample

sizes for groups of PSUs which will produce the lowest variance given the available budget for travel and data collection. The model was generalized by allowing the number of non-certainty PSUs in each non-self representing index area to be a variable that the optimization program could optimize over. This created the need for an additional constraint though as the number of non-certainty PSUs was determined by the total number of HSEs minus the number of HSEs used by the certainty PSUs.

In addition, the relative importances for each index area and group of items had to be recalculated for each scenario. The populations used for calculating the population relative importances were from the 2000 Census. The cost weights used for calculating the relative importance of groups of items were from the 1999 Consumer Expenditure survey. An index area as used in this paper is either a certainty city or a Census region by size class combination. There are four Census regions: Northeast, Midwest, South, and West. There are two size classes corresponding to metropolitan and micropolitan CBSAs. Note that some micropolitan CBSAs are part of the current certainty cities and thus their population should be included with the certainty city and not with the non-self representing index area covering micropolitan CBSAs in the Census region in which the PSU resides..

Some additional options were explored. Even though we currently allocate one HSE to each non-certainty PSU, there was interest in what would happen if two HSEs were allocated to each non-certainty PSU. It would be expected to roughly halve the number of non-certainty PSUs, but the effect on variance was less obvious. Also, there was concern that the grossly uneven relative importances of the index areas may have a negative impact on sample allocation and on the variance of the all U.S. – all items CPI-U. Thus an option was explored

where the largest Census region, the South, was broken apart using Census divisions. The South was divided into two index areas, one being the South Atlantic division and the other index area being composed of the East South Central and West South Central divisions. New variance components for the optimization model were calculated for the new index areas.

The optimization model yielded a result with non-integer numbers of PSUs in each non-self representing index area. These values were rounded to even integers in such a way that the total number of HSEs added up correctly. The optimization model was then rerun using these fixed numbers of PSUs to provide results that could be compared with results from other scenarios. The information was used in determining what the set of certainty PSUs would be.

The list of certainty PSUs is not yet public information and can't be included in this paper. Some of the results that were found can be discussed. In comparing the allocation of one vs. two HSEs to each non-certainty PSU, it was found that allocating two HSEs to each non-certainty PSU increased the modeled standard error of six month CPI change for C&S by an average of 13.6% across the scenarios. This was primarily due to the large contribution of the between PSU component of variance in non-self representing index areas. This was surprising given that the PSU components of variance are so small compared to other components of variance. However the much smaller divisor of the PSU component of variance as compared to other components allowed it to have a greater contribution to the total variance. In all cases the PSU component of variance ended up contributing more than 50% of the total variance for all of the index areas representing metropolitan CBSAs.

Dividing the South based on Census divisions also ended up increasing the total variance. It appears based upon the model used that it is preferable to have fewer and larger index areas with larger PSU samples than to have a larger number of smaller index areas. This is again a result of the large contribution of the between PSU component of variance of non-self representing index areas.

Once the decision was made on a set of certainty PSUs, the number of PSUs in each non-self representing index area was also determined based on the output of the optimization program from the chosen scenario. The chosen design did shift towards having more PSUs in the West region and slightly fewer elsewhere. There are more of what are called C-size PSUs as the population they cover has grown greatly in relative importance between 1990 and 2000. For the 1998 CPI Revision sample, the C PSUs were the urban part of areas outside of metropolitan statistical areas. The C PSUs now represent the micropolitan CBSA population, excluding those CBSAs which are part of a certainty PSU. Having the CPI-U population be the total population in CBSAs resulted in an increase in the total percent of the U.S. population covered by the CPI-U.

Stratifying Non-Certainty PSUs

Non-certainty PSUs are grouped together into strata and one PSU is selected from each stratum. (see Dipppo et al) It is desirable that the PSUs within a stratum be homogeneous. The first task was to determine by what measure the PSUs should be homogeneous.

It the early 1990's, work was done on modeling CPI-U change for certainty PSUs by variables we had available from Census as well

as geographic variables. None of these models were especially promising. However, for the 1998 CPI Revision, a four variable model using normalized latitude, normalized longitude, normalized latitude squared, and percent urban was chosen for use in three out of four Census regions and a model consisting of seven Census variables was chosen for the South region. Once a model was chosen, the strata were formed so as to be as homogeneous as possible with respect to these variables, subject to the restriction that strata should have roughly equal population. (see Williams et al)

This research was updated by examining the predictive power of these models for more recent time periods as well as examining their value in modeling CPI-U change for non-self representing PSUs and for modeling changes in the housing index. The chosen models have performed worse since they were originally researched and no other really good models have been found. Thus the chosen model this time was simply the four variable model from before with normalized longitude squared included for the purpose of symmetry.

Given the relatively weak predictive power of the chosen model, two other options were also examined: Using no stratification and a purely geographic stratification.

With no stratification, the PSUs would be drawn from each region by size class without replacement and with probability proportional to expenditure. This was done for simulation purposes with SAS PROC SURVEY SELECT.

The purely geographic stratification was based on Peano ordering the PSUs based on the median latitude and longitude of the centroids of the counties composing the PSUs. Examples of Peano curves can be found at

<http://www.contrib.andrew.cmu.edu/~malin/java/PeanoHilbert.html>. The Peano curve for a $2^N \times 2^N$ grid is based on a recursive N-shaped pattern. In each region by size class combination, the points representing the PSUs were placed on a $2^{20} \times 2^{20}$ grid. The calculation of an ordering value is based on interleaving the digits of the binary representations of the coordinates of the PSUs. Once the PSUs are ordered, the ordered list of PSUs in each region by size class is cut into the appropriate number of strata. The cut points are made so that the population in each stratum is roughly the same. It was also attempted to make the cut points such that when there was a large jump in the calculated ordering value between two points then the two points would fall in different strata. This purely geographic stratification ended up producing strata which looked like rectangular stripes.

In order to cluster PSUs to be similar according to the five variable model discussed above, a program using a hill climbing algorithm by Friedman and Rubin was used. This program first rescales all of the variables so that they are of roughly equal importance. It does this by calculating an unstratified population weighted sum of squares for each of the variables and then multiplies the values of the variables by ten divided by the square root of the sum of squares:

$$V'_{i,j} = V_{i,j} * \frac{10}{\sqrt{\text{Total Population} * \sum_j \text{PSU}pop_j * (V_{i,j} - \bar{V}_{i,j})^2}}$$

where

$V_{i,j}$ is the value of the i th variable for the j th PSU

$\text{PSU}pop_j$ is the population of the j th PSU

$$\text{Total Population} = \sum_j \text{PSU}pop_j$$

$$\bar{V}_{i,j} = \sum_j \frac{\text{PSU}pop_j}{\text{Total Population}} * V_{i,j}$$

The program then attempts to minimize the stratified total sums of squares

$$\sum_{i,s} Stratumpop_s * \sum_{j \in S} PSUpop_j * \left(V'_{i,j} - \sum_{k \in S} \frac{St}{k} \right)$$

given the total number of strata, which is an input to the program. This program repeats the minimization procedure to form strata in each Census region by size class. The program is constrained on the size of the strata, and these constraints were estimated using the minimum and maximum stratum populations from the geographic stratification and adjusting them by 10%.

Keyfitzing to increase overlap

Given our budgetary limitations, it is generally desirable to keep as many of our current PSUs in the next sample as possible.

The first step was to determine what is meant by an overlap PSU. Given the considerable changes in definitions of the PSUs it is possible that part of a PSU might currently be in the CPI sample but not other parts. The preliminary definition was that 30% of the counties or 30% of the 2000 population of a PSU currently be covered by the CPI sample. This was complicated by the fact that counties are composed of Minor Civil Divisions (MCDs) in the Northeast region. Current CPI PSUs in the Northeast are defined at the MCD level, while the new PSUs are defined at the county level. It was decided that a county composed of MCDs was overlap if at least 5% of its 2000 population was overlap. A PSU composed of MCDs is considered overlap as long as 30% of the counties are overlap and at least one of those counties has at least 30% of its 2000 population being overlap based on MCDs.

The inherited Keyfitzing procedure attempts to increase the likelihood of selecting PSUs which are overlap, or which have a greater relative importance in 2000 than in 1990. Some changes in the program had to be made due to the massive redefinition of PSUs. The Keyfitzing procedure operates at the level of the intersection of a new stratum with a stratum for the 1998 CPI Revision PSU sample. Due to redefinitions, there are many cases where only part of a new PSU lies within one of these intersections. Thus the PSUs were broken in pieces for the purpose of Keyfitzing and then the pieces were added together to give the total new probability of selection of a PSU.

The procedure works as follows:

For each Region X City Size X New Stratum_i X Old Stratum_j calculate the new probability of the PSU k or the part of PSU k being selected:

$$Newoldpr_{i,j,k} = \frac{newprob_k}{\sum_{l \in new_i \cap old_j} newprob_l}$$

where $newprob_k$ is the probability of selection of the intersection of PSU k with new stratum i and old stratum j.

There are several possible cases:

a) The intersection is empty so there are no PSUs to consider

b) The intersection is a single PSU k. Then the Keyfitted probability is $Keyfitz_k = newprob_k$

c) There is no PSU in the intersection which was selected in the old sample:

For each PSU k in the intersection assign the Keyfitz probability as

If $Newoldpr_{i,j,k} \leq oldprob_k$ then
 $Keyfitz_k = 0$

If $Newoldpr_{i,j,k} > oldprob_k$ then
 $Keyfitz_k = \frac{Newoldpr_{i,j,k} - oldprob_k}{\sum_{l \in new_i \cap old_j} \max(Newoldpr_{i,j,l} - oldprob_l)}$

* $\sum_{l \in new_i \cap old_j} newprob_l$

Here $oldprob_k$ is the probability of selection of PSU k intersected with new stratum i and old stratum j based on 1990 populations.

d) A PSU s was selected in the old sample and at least partially resides in the intersection:

If $Newoldpr_{i,j,s} \geq oldprob_s$ then
 $Keyfitz_s = \sum_{l \in new_i \cap old_j} newprob_l$

$Keyfitz_k = 0$ for all other PSUs k within the intersection.

Here the new and old probabilities are based on the old PSU definition for PSU s intersected with new stratum i and old stratum j. The Keyfitz probability for new PSUs within the intersection of new stratum i and old stratum j is calculated by determining the percentage of 2000 population of the old PSU s resides within each of the new PSUs.

e) A PSU s was selected in the old sample and at least partially resides in the intersection:

If $Newoldpr_{i,j,s} < oldprob_s$ then
 $Keyfitz_s = \frac{newprob_s}{oldprob_s} \sum_{l \in new_i \cap old_j} newprob_l$

If k is a PSU in the intersection other than s, then

if $Newoldpr_{i,j,k} \leq oldprob_k$ then
 $Keyfitz_k = 0$

else if $Newoldpr_{i,j,k} > oldprob_k$ then

$$Keyfitz_k = \left(\sum_{l \in new_i \cap old_j} newprob_l \right) * \left(1 - \frac{newprob_s}{oldprob_s} \right) * \frac{Newoldpr_{i,j,k} - oldprob_k}{\sum_{l \in new_i \cap old_j} \max(Newoldpr_{i,j,l} - oldprob_l, 0)}$$

After this procedure has been done for each intersection of new and old strata then the PSUs are reaggreated and their total probabilities of selection are determined.

The selection of a stratification was made on the basis of the total expected number of overlap PSUs. It turned out that the stratifications with the highest overlap were from the clustering procedure using normalized latitude, normalized longitude, normalized latitude squared, normalized longitude squared, and percent of population which is urban. As the clustering procedure had been run multiple times, there was usually more than one stratification to choose from in each Census region by size class. It turned out that having a lower total sums of squares did not equate with having higher expected overlap.

The following table summarizes the expected number of overlap PSUs for the various options examined, both pre- and post-Keyfitzing:

Region – City size	R C	#over lap PSUs No stratificati on	# over lap PSUs P eano ordering	#over lap PSUs cluste ring program
100	X	1.0 7	1 .01	0.98
200	X	5.0 0	4 .56	4.30
300	X	5.0 6	5 .01	4.96
499	X	1.4 5	1 .37	1.40
100	C	0.1 0	0 .05	0.05
200	C	0.2 4	0 .23	0.22
300	C	0.2 7	0 .30	0.30
400	C	0.3 5	0 .34	0.34
000	X	12. 58	1 1.95	11.64

499		
100	C	0.05
200	C	0.23
300	C	0.30
400	C	0.34
000	X	18.53
		23.27

Controlled selection of PSUs

It is hoped that the number of overlap PSUs selected is not much less than the expected number of overlap PSUs. Thus a procedure called controlled selection was used. A program used to do the controlled selection for the 1998 CPI Revision PSU sample could not be successfully compiled and run in our current computing environment. An alternative called PC Consel (see Lin) was investigated. We had some success with this program, however in the South region it would not give a solution as apparently no exact solution to the controlled selection problem exists. Thus a new SAS IML program was written in order to handle the controlled selection problem.

The following is a description of the controlled selection problem:

Create a 3-dimensional grid of stratum x state x overlap status. Sum the probabilities of selection of the PSUs in each cell. A pattern describes an entire sample. In each cell it has either a zero (select zero PSUs from this cell) or

Region – City size	R C	#over lap PSUs Peano ordering after Keyfitzing	#over lap PSUs cluste ring program after Keyfitzing
100	X	2.43	2.82
200	X	6.56	8.40
300	X	6.44	7.59
	X	3.10	4.46

one (select one PSU from this cell). The controlled selection problem is to find a set of patterns $P_i, i=1, \dots, n$ with probabilities of selection p_i such that

$$\sum_{i=1}^n p_i * P_i(x, y, z) = C(x, y, z), \quad \text{where}$$

$P_i(x, y, z)$ is the value of zero or one for the i th pattern for stratum x , state y , and overlap status z and $C(x, y, z)$ is the sum of probabilities of selection of PSUs in the cell for stratum x , state y , and overlap status z .

In addition there are constraints with respect to the number of PSUs selected per state and per overlap status. These constraints are imposed on each individual pattern. Let $S_i = \sum_x \sum_z C(x, i, z)$ be the total probability of PSUs in state i . Let $\underline{S}_i = \text{floor}(S_i)$ be the integer part of S_i . Then each pattern must contain either \underline{S}_i or $\underline{S}_i + 1$ PSUs in state i . The sum of probabilities of patterns having \underline{S}_i PSUs is $1 - (S_i - \underline{S}_i)$ and the sum of probabilities of patterns having $\underline{S}_i + 1$ PSUs is $S_i - \underline{S}_i$.

Let $O = \sum_x \sum_y C(x, y, 1)$ be the sum of probabilities of selection of overlap PSUs across all strata and states.

Let $\underline{O} = \text{floor}(O)$ be the integer part of O . Then each pattern must select \underline{O} or $\underline{O} + 1$ overlap PSUs. The sum of probabilities of patterns with \underline{O} overlap PSUs is $1 - (O - \underline{O})$ and the sum of probabilities of patterns with $\underline{O} + 1$ overlap PSUs is $O - \underline{O}$.

The above constraints on the set of patterns comprises the controlled selection problem. Once this problem is solved, a pattern is selected based on the probabilities of the patterns. If there is more than one PSU corresponding to a cell with a value of one, then

a single PSU is selected with probability proportional to its probability of selection within its stratum.

Note that there isn't necessarily a solution for the controlled selection problem. If there is no exact solution, then it is desirable to have a partial set of patterns $P_i, i=1, \dots, n$ which have a sum of probabilities as close to one as possible.

The program randomly generates patterns by selecting a value of zero or one in each cell of the pattern using the probability in that cell. The program then verifies that the pattern meets the state and overlap constraints. If the pattern violates any constraints then the pattern is discarded and a new pattern is generated. If the pattern meets the state and overlap constraints then the pattern is kept and it is assigned a probability. The probability assigned to the pattern is the smallest remaining probability in any cell where a PSU was selected or the smallest remaining probability of the state and overlap controls met:

Let

$$\min \text{cell}_i = \min_{x,y,z} C(x, y, z) * P_i(x, y, z)$$

For each state i , the associated probability with the constraint is $1 - (S_i - \underline{S}_i)$ if \underline{S}_i PSUs are selected and $S_i - \underline{S}_i$ if $\underline{S}_i + 1$.

For the overlap constraint, the associated probability is $1 - (O - \underline{O})$ if \underline{O} overlap PSUs are selected in the pattern and $O - \underline{O}$ if $\underline{O} + 1$ overlap PSUs are selected.

The probability assigned to the pattern is the minimum of the cell probabilities, the state constraint probabilities, and the overlap constraint probability.

Once the pattern has a probability, that probability is deducted from each cell where a PSU was selected as well as from the state and overlap constraints met. For example, if the pattern probability is 0.2 and the number of PSUs in a state with 2.4 expected PSUs is 2, then the 0.6 probability initially assigned to selecting 2 instead of 3 PSUs in that state would be reduced to 0.4.

The new problem with the probabilities subtracted now goes through the same procedure until all probability is exhausted.

The way the patterns are constructed and the probabilities assigned, the sum of probabilities of patterns where a given PSU is selected will add up to the probability of the given PSU being selected. In addition, the probabilities associated with the state and overlap constraints will add up properly.

References:

Dippo, Cathryn S., and Jacobs, Curtis A., "Area Sample Redesign for the Consumer Price Index," *Proceedings of the Survey Research Methods Section*, American Statistical Association, 1983, 118-123.

Leaver, Sylvia G., Johnson, William, Shoemaker, Owen, and Benson, Thomas S., (1999) "Sample Redesign for the Introduction of the Telephone Point of Purchase Survey Frames In the Commodities and Services Component of the U.S. Consumer Price Index ," *Proceedings of the Section on Government Statistics, American Statistical Association*, 1999, 292-297.

Lin, Ting-Kwong, "Some Improvements on an Algorithm for Controlled Selection," *Proceedings of the Survey Research Methods Section*, American Statistical Association, 1992, 407-410.

Williams, J.L., Brown, E.F., Zion, G.R., "The Challenge of Redesigning the Consumer Price Index Area Sample," *Proceedings of the Survey Research Methods Section* , American Statistical Association (Vol. 1), 1993, 200-205.

**EVALUATION OF VENDOR LISTS
USING CENSUS DATA**

Submitted by:

Westat
1650 Research Blvd.
Rockville, MD 20850

September 25, 2002



TABLE OF CONTENTS

<u>Section</u>		<u>Page</u>
1	INTRODUCTION.....	1
2	DATA.....	2
3	EVALUATION OF LISTS AT BLOCK LEVEL.....	3
	3.1 Coverage of Lists.....	3
	3.2 Accuracy of Renter Distribution.....	5
	3.3 Assessment Using CPI Data.....	16
4	SUMMARY.....	20

List of Appendices

<u>Table</u>		<u>Page</u>
A	DETAILED TABLES.....	A-1

List of Tables

<u>Table</u>		<u>Page</u>
1	Number of housing units from SF1, MSG, and Dunhill, by county.....	4
2	Percent of census blocks not found in lists, by SF1 characteristics.....	5
3	Percent renter using two different coding schemes for the MSG file, by county.....	6
4	Percent renter for the SF1, MSG, and Dunhill files, by county.....	6
5	Percent of blocks in MSG and SF1, by percent renter.....	7
6	Percent of blocks in Dunhill and SF1, by percent renter.....	7
7	Percent of blocks in MSG and SF1, by percent renter and state.....	8
8	Percent of blocks in Dunhill and SF1, by percent renter and state.....	9
9	Percent of blocks in MSG and SF1, by percent renter and county.....	9
10	Percent of blocks in Dunhill and SF1, by percent renter and county.....	10

TABLE OF CONTENTS (CONTINUED)

List of Tables (Continued)

<u>Table</u>		<u>Page</u>
11	Percent of blocks in MSG and SF1, by percent renter and block size.....	12
12	Percent of blocks in Dunhill and SF1, by percent renter and block size.....	13
13	Ratios of the number of occupied housing units in a block for MSG and Dunhill to SFI, by geography.....	15
14	Ratios of the number of rented housing units in a block for MSG and Dunhill, by geography.....	15
15	Ratios of the number of owned housing units in a block for MSG and Dunhill, by geography.....	16
16	Number of blocks in CPI and SF1, by percent renter and CPI sample size.....	17
17	Percent of blocks in CPI and SF1, by percent renter and CPI sample size.....	18
18	Percent of blocks in CPI and MSG, by percent renter and CPI sample size....	19
19	Percent of blocks in CPI and Dunhill, by percent renter and CPI sample size.	19

1. Introduction

The Office of Prices and Living Conditions of the Bureau of Labor Statistics (BLS) is exploring the use of purchased lists of addresses to enhance or replace the in-person listing and screening processes used to identify renters in the Consumer Price Index (CPI) Housing Survey. This report describes the evaluation of the lists based on aggregates of census block level data. Another task and report will evaluate the lists at the individual housing unit level based on data collected from a sample of housing units.

To conduct the research, we contacted three vendors to obtain lists of housing units for selected counties in Baltimore and Richmond Metropolitan Statistical Areas. The counties included in the Baltimore MSA were Baltimore County, Howard County, and Queen Anne's County. In the Richmond MSA, the two counties included were Hanover County and Henrico County.

Westat purchased lists from Marketing Systems Group (MSG) and Dunhill International and contacted Experian. Experian could not provide data at this time⁴ because the data on housing units in its files do not contain 2000 Decennial Census block identifiers needed for matching the data to the census files. The costs of obtaining the lists were \$15,000 for the MSG list and \$11,000 for the Dunhill list. Both the MSG and Dunhill lists use U.S. Postal delivery addresses as the base and then append additional data from other sources. The prime data source for the MSG list is Info USA and Dunhill uses Knowledgebase. BLS also provided us with block-level data from the CPI Housing Survey in the targeted counties.

This report gives census block-level summaries that compare the lists from the vendors with the corresponding block-level data from the 2000 Decennial Census. Comparisons are made for the total number of housing units, the number and proportion of rented housing units, and the number and proportion of owned housing units. These comparisons are given by state, county, percent renter, and by block size using the data from the 2000 Decennial Census. In addition to summary information in this report, a detailed dataset containing block-level data from each list and from the census is being submitted. This report also compares the block level information from the lists with CPI sample information. A limitation of the comparisons to the CPI data is that the data from the CPI are only available for a sample of blocks in the MSAs and the sample size in some of the blocks is small.

As noted above, the second task will involve selecting a sample of households from the lists and comparing the data for the housing unit to data collected from the sampled housing units. An

⁴ Experian indicated they may update their files at some later date.

important goal of this task is to examine the accuracy of tenure (own/rent) data from the lists. This task will be covered in a separate report.

2. Data

Both MSG and Dunhill were asked to provide a list for the specified counties in the two MSAs. The list was to include data identifying the state, county, tract, tract subgroup, block group, block, address, name (where available), telephone number (where available), and tenure (where available). The list from Dunhill had more than 5,000 cases with missing data for the tract subgroup field. We filled these blanks with zeros and Dunhill verified with their data source that this was appropriate.

The list must contain census blocks because the records are matched to block-level data from 2000 Decennial Census Summary File 1 (SF1). A total of 13,262 blocks with at least one occupied housing unit⁵ were identified and extracted from the SF1 in the five specified counties.

The MSG file contained 557,559 records from the five specified counties. Each record corresponds to a housing unit. Of these records 350,845 (63%) have a telephone number available. The tenure variable on the MSG list has values ranging from 0 to 9. The value 0 means the housing unit is a 'definite' renter and the value 9 means the unit is a 'definite' owner. As the values go from 1 to 8 the likelihood of the unit being a rental decreases.

The Dunhill file contained 471,868 records from the five specified counties. As with the MSG file, each record corresponds to a housing unit. Of these records 272,405 (58%) have a telephone number available. The tenure variable from Dunhill is a dichotomy that indicates the unit is either owned or rented.

Both the MSG and the Dunhill files do classify the housing units by census block, but this data is inferred from the address using various geographic coding techniques. For example, MSG indicated that the block was assigned based on the ZIP+4 Code. Thus, it is possible, and even likely in some cases, that the data for a housing unit is classified as being in a particular census block but it actually falls in another block. This report cannot assess this type of error because we do not have access to housing unit level data from the 2000 census.

⁵ Some census blocks contain no occupied housing units and those blocks were excluded from the analysis. Including blocks with no occupied housing units there are 16, 549 blocks in the SF1.

The CPI Housing Survey data is for a sample of 2000 Census blocks in the five counties. BLS created a block-level summary with the number of housing units sampled in the block, the number of these that are owned and the number of these that are rented. The file contains 285 blocks that summarize the data of 2,799 sampled housing units. Of the 285 blocks, we could not match five of the blocks to the SF1 data.

3. Evaluation of Lists at Block Level

The first issue addressed is the coverage of the MSG and Dunhill lists. By coverage, we mean the number of housing units identified in the list as compared to the number of occupied housing units counted in the SF1 file. We assume the SF1 file is complete and accurate for this purpose.⁶

3.1 Coverage of Lists

The first step of the process was to determine if the two lists could be matched to a corresponding census block from the SF1 file in the selected counties. After summarizing the lists at the block-level and dealing with the blank data in the Dunhill file, the housing units from the two files were matched to the SF1 blocks. The MSG file had records in 809 blocks that were not in the SF1 file (in the selected counties) and the Dunhill file contained records in 1,234 blocks that were not in the SF1 file.

Table 1 gives the number of occupied housing units in the SF1 file and in the lists from MSG and Dunhill. The numbers are given for all the data from the lists, and just for the blocks that match to a census block on the SF1 file. The records from the lists that do not match to the SF1 blocks were eliminated from subsequent analysis and are not further discussed unless specifically noted.

Even after restricting the analysis to the matching blocks, the MSG list still has either about the same or more housing units than the SF1 file for each of the five counties. The list from Dunhill has fewer housing units than both the MSG list and the SF1 file in each county.

⁶ Clearly, some changes have happened between 2000 and 2002 but we have no means of evaluating these changes.

Table 1. Number of housing units from SF1, MSG, and Dunhill, by county

Geography	Number of housing units	SF1	MSG	Dunhill
Total	Total	544,477	557,559	471,868
	In matching blocks		548,929	456,579
County				
	Baltimore			
	Total	299,877	303,191	258,377
	In matching blocks		299,869	252,954
Howard	Total	90,043	96,146	81,435
	In matching blocks		93,568	78,551
Anne's Queen	Total	15,315	15,308	14,376
	In matching blocks		14,642	13,305
Hanover	Total	31,121	32,981	30,084
	In matching blocks		32,539	27,360
Henrico	Total	108,121	109,933	87,596
	In matching blocks		108,311	84,409

NOTE: The SF1 counts are for occupied housing units.

We also examined the files to determine if there were some blocks with occupied housing units in the SF1 without any corresponding data from the lists. Out of the 13,262 census blocks in the SF1 file in the five counties, the list from MSG had data from 12,329 blocks and the list from Dunhill had data from 12,176 blocks. To investigate the census blocks without corresponding data from the lists, we examined the characteristics of the SF1 blocks that were not matched to the lists. Table 2 gives the percentage distribution of the 933 SF1 blocks that were not found in the MSG file and the 1,086 SF1 blocks that were not found in the Dunhill file by the characteristic of the block in the SF1 file. The distributions for MSG and Dunhill are relatively consistent by the SF1 characteristic. Two variables that are highly related to the missingness are block size and percent renter. Over 20 percent of the SF1 blocks with 10 or fewer occupied housing units do not have a corresponding block-level record in either the

MSG or Dunhill file. The blocks with zero percent renter in the SF1 file are missing at a high rate for both MSG and Dunhill, as are the blocks with over 40 percent renters. The missingness also varies considerably by county.

The comparisons in terms of coverage at this level indicate that the gross coverage rate for the MSG file is better than that of the Dunhill file. In fact, the MSG counts indicate more housing units than are in the SF1 data file. The Dunhill coverage rate is good, but less than the SF1 counts.

The tables in the rest of the report are typically given as either percentages or ratios for the numbers in the blocks that are in both the SF1 and the list (MSG or Dunhill). The appendix contains the counts of the data. The counts for the SF1 blocks that are not in the list are also given in these tables. The percentages for all the SF1 blocks can be computed from these tables, although these percentages are not the most informative ones for evaluating the accuracy of the list data.

Table 2. Percent of census blocks not found in lists, by SF1 characteristics

SF1 block characteristic	MSG missing blocks	Dunhill missing blocks
Total	7.0%	8.2%
State		
Maryland	7.2%	8.5%
Virginia	6.6%	7.5%
County		
Baltimore	5.5%	6.0%
Howard	9.0%	10.8%
Queen Anne's	22.4%	16.8%
Hanover	16.1%	15.0%
Henrico	4.6%	3.7%
Block size (occupied units)		
$0 < s \leq 10$	20.3%	22.9%
$10 < s \leq 30$	3.0%	3.7%
31 or more	1.5%	2.2%
Percent renter		
$r=0$	12.8%	14.3%
$0 < r \leq 10$	1.2%	1.9%
$10 < r \leq 20$	3.0%	3.8%
$20 < r \leq 30$	4.4%	5.6%

30<r≤40	4.6%	6.9%
r>40	12.6%	14.1%

3.2 Accuracy of Renter Distribution

As noted previously, the MSG file contains a variable with codes ranging from 0 to 9 that could be used to classify the tenure status of the unit. To summarize the data, we computed two binary own/rent variables from these data. The first variable classified units with codes of 0 to 6 as renters and 7 to 9 as owners. The second variable classified units with codes of 0 to 5 as renters and 6 to 9 as owners. Table 3 shows the percent renters for each county using both variables and the SF1 percent renter using all the data from the MSG file prior to matching. Because the second scheme (0 to 5 renters/6 to 9 owners) gives a closer match to the SF1 percent, we use this scheme throughout the report unless otherwise noted.

Table 3. Percent renter using two different coding schemes for the MSG file, by county

County	SF1	MSG (0-6/7-9)	MSG (0-5/6-9)
Baltimore	32.5 %	34.3%	31.8%
Howard	26.2 %	29.6%	27.0%
Queen Anne's	16.6 %	21.8%	15.1%
Hanover	15.7 %	17.8%	13.7%
Henrico	34.3 %	37.7%	34.3%

NOTE: These percentages use all the MSG data, not just those in the matching blocks.

Table 4 gives the distribution of the percent renter for the SF1, MSG and Dunhill files for all records (not just the matching blocks). The MSG distribution uses the second scheme (0 to 5 renters/6 to 9 owners). This table shows that the percent renter from the MSG list is closer to the SF1 percent renter than the Dunhill list for all but Hanover County. Even within county the difference between percentages are not large. The percent renter from the Dunhill list is more than 6 percentage points different from the

SF1 percent in three of the five counties, while the MSG percent renter never differs from the SF1 percent by more than 2 percentage points. This analysis suggests the MSG list might match more closely than the Dunhill list on this characteristic, but further investigation using the matching blocks is required and is given below.

Table 4. Percent renter for the SF1, MSG, and Dunhill files, by county

County	SF1	MSG	Dunhill
Baltimore	32.5%	31.8%	26.3%
Howard	26.2%	27.0%	24.0%
Queen Anne's	16.6%	15.1%	24.2%
Hanover	15.7%	13.7%	17.1%
Henrico	34.3%	34.3%	23.5%

NOTE: These percentages use all the MSG data, not just those in the matching blocks. The MSG percent rental uses codes 0 to 5.

The tables below compare the distribution of the renter occupied housing units in the lists to the SF1 distribution for the matching blocks (the 12,329 MSG blocks that match to the SF1 and the 12,176 Dunhill blocks that match to the SF1). Table 5 gives the percent distribution for the matching MSG blocks categorized by the SF1 percent renter distribution. For example, 51 percent of the blocks that are classified by the SF1 as having no renters ($r=0$) are blocks that are also classified into this category in the MSG file. The diagonal elements where the SF1 and the MSG categories are identical are in bold. The MSG designation for blocks with greater than 40 percent renter is 72 percent, the highest in the table. However, this is partially due to the categorization scheme. For example, if only two categories were used, then 95 percent of the blocks categorized as 40 percent or less renters would be in the same category using the MSG data.

Table 6 gives the same distribution for the Dunhill file compared to the SF1 data. The distribution for the Dunhill file is similar to that of the MSG file. If only two categories were used to summarize these data, then 97 percent of the blocks categorized as 40 percent or less renters would be in the same category using the Dunhill data. The main difference between the MSG and Dunhill percent distributions is for the extreme categories. For the Census blocks with 0 to 10 percent renter, the list from Dunhill has a more blocks in the same category than the list from MSG. On the other hand, the list from MSG has a higher percentage for categories with more than 20 percent renters.

Table 5. Percent of blocks in MSG and SF1, by percent renter

SF1	MSG						Total
	r=0	0<r≤10	10<r≤20	20<r≤30	30<r≤40	r>40	
r=0	5 1%	2 4%	1 3%	4 %	3 %	5 %	1 00%
0<r≤10	2 2%	4 9%	1 8%	6 %	2 %	2 %	1 00%
10<r≤20	1 7%	3 0%	2 7%	1 5%	6 %	4 %	1 00%
20<r≤30	1 5%	1 7%	2 5%	2 2%	1 1%	9 %	1 00%
30<r≤40	1 5%	6 %	2 1%	1 9%	1 8%	2 1%	1 00%
r>40	1 1%	2 %	4 %	5 %	7 %	7 2%	1 00%
Total	2 7%	2 8%	1 7%	9 %	5 %	1 5%	1 00%

NOTES: Table may not add to 100 percent due to rounding.

The variable 'r' is the percentage of the occupied housing units in the block that are rental.

Table 6. Percent of blocks in Dunhill and SF1, by percent renter

SF1	Dunhill						Total
	r=0	0<r≤10	10<r≤20	20<r≤30	30<r≤40	r>40	
r=0	4 9%	2 8%	1 4%	4 %	2 %	3 %	100%
0<r≤10	1 9%	5 8%	1 7%	3 %	1 %	2 %	100%
10<r≤20	2 1%	3 7%	2 9%	8 %	3 %	3 %	101%
20<r≤30	2 0%	2 4%	3 4%	1 2%	5 %	4 %	99%

≤40	30<r	2	1	2	1	9	1	99%
		0%	3%	7%	9%	%	1%	
	r>40	1	3	8	9	1	5	100%
		2%	%	%	%	0%	8%	
Total		2	3	1	6	3	1	99%
		7%	3%	9%	%	%	1%	

NOTES: Table may not add to 100 percent due to rounding.

The variable 'r' is the percentage of the occupied housing units in the block that are rental.

The next tables give the same data broken down by state and county. Table 7 is the percent renter distribution by state for the MSG file and Table 8 is the percent renter distribution by state for the Dunhill file. The distributions are very similar across state for both the MSG and Dunhill files, with no remarkable differences. Table 9 gives the percent renter distribution by county for the MSG file and Table 10 is the percent renter distribution by county for the Dunhill file. As expected, the distributions by renter status from both sources are more variable at the county level than the state level. Two counties, Queen Anne's and Hanover, are very different from the other counties. These two counties have much lower percentages in the over 40 percent renter category for both MSG and Dunhill than observed in the other counties. For example, Queen Anne's county has only 26 percent in the diagonal for the over 40 percent category for MSG while the average for this category across all the counties is 72 percent (the corresponding percentage for the Dunhill file is 21 percent compared to the overall average of 58 percent). We are not aware of any reason for the problems in these two counties, but it does suggest that a wider range of counties may have to be examined to assess the quality of the lists.

Table 7. Percent of blocks in MSG and SF1, by percent renter and state

SF1	MSG						Total
	r=0	0<r≤10	10<r≤20	20<r≤30	30<r≤40	r>40	
MD							
r=0	5	2	1	4	3	5	100
	1%	5%	2%	%	%	%	%
0	0<r≤1	2	5	1	6	2	100
	2%	0%	8%	%	%	%	%
0	0<r≤1	1	3	2	1	6	100
	7%	1%	7%	5%	%	%	%

30	20<r≤	1	1	2	2	1	8	100
		7%	7%	5%	2%	1%	%	%
40	30<r≤	1	7	1	1	1	2	101
		7%	%	9%	8%	7%	3%	%
	r>40	1	2	4	5	6	7	100
		1%	%	%	%	%	2%	%
Total		2	2	1	8	5	1	101
		7%	9%	7%	%	%	5%	%
VA								
	r=0	5	2	1	5	3	4	100
		2%	3%	3%	%	%	%	%
0	0<r≤1	2	4	1	7	3	2	100
		4%	5%	9%	%	%	%	%
0	0<r≤1	1	2	2	1	7	5	100
		7%	8%	8%	5%	%	%	%
30	20<r≤	1	1	2	2	1	9	100
		3%	7%	6%	4%	1%	%	%
40	30<r≤	1	5	2	2	2	1	100
		1%	%	5%	1%	0%	8%	%
	r>40	1	2	4	5	8	7	99
		0%	%	%	%	%	0%	%
Total		2	2	1	1	6	1	100
		7%	5%	8%	0%	%	4%	%

NOTES: Table may not add to 100 percent due to rounding.

The variable 'r' is the percentage of the occupied housing units in the block that are rental.

Table 8. Percent of blocks in Dunhill and SF1, by percent renter and state

SF1	Dunhill						Total
	r=0	0<r≤10	10<r≤20	20<r≤30	30<r≤40	r>40	
MD							
r=0	4	2	1	4	2	4	100
	7%	8%	5%	%	%	%	%
0<r≤	1	6	1	4	1	2	101

10		7%	0%	7%	%	%	%	%
	0<r≤	1	3	3	8	3	3	99%
10		8%	6%	1%	%	%	%	
	20<r	1	2	3	1	6	5	100
≤30		8%	0%	6%	5%	%	%	%
	30<r	1	1	2	2	1	1	100
≤40		8%	1%	4%	3%	2%	2%	%
	r>40	1	2	6	9	1	6	100
		1%	%	%	%	0%	2%	%
Total		2	3	1	7	4	1	100
		4%	4%	9%	%	%	2%	%
	VA							
	r=0	5	2	1	3	1	2	99%
		3%	7%	3%	%	%	%	
	0<r≤	2	5	1	3	1	1	99%
10		6%	2%	6%	%	%	%	
	0<r≤	2	3	2	7	1	1	98%
10		7%	7%	5%	%	%	%	
	20<r	2	3	3	8	4	3	100
≤30		3%	1%	1%	%	%	%	%
	30<r	2	1	3	1	5	9	100
≤40		4%	7%	3%	2%	%	%	%
	r>40	1	4	1	9	1	5	99%
		4%	%	1%	%	1%	0%	
Total		3	3	1	6	3	9	101
		2%	2%	9%	%	%	%	%

NOTES: Table may not add to 100 percent due to rounding.

The variable 'r' is the percentage of the occupied housing units in the block that are rental.

Table 9. Percent of blocks in MSG and SF1, by percent renter and county

SF1	MSG						To tal
	r= 0	0 <r≤10	1 0<r≤20	2 0<r≤30	3 0<r≤40	r> 40	
Baltimore							

County									
$r=0$	4	2	1	4	2	4	99		
	9%	7%	3%	%	%	%	%		
$0 < r \leq 10$	2	4	2	7	2	1	10		
	1%	9%	0%	%	%	%	0%		
$0 < r \leq 10$	1	3	2	1	6	3	10		
	6%	2%	7%	6%	%	%	0%		
$20 < r \leq 30$	1	1	2	2	1	7	10		
	6%	8%	5%	3%	1%	%	0%		
$30 < r \leq 40$	1	9	2	1	1	2	10		
	7%	%	1%	7%	6%	1%	1%		
$r > 40$	7	2	4	3	6	7	10		
	%	%	%	%	%	8%	0%		
Total	2	2	1	9	5	1	10		
	5%	9%	7%	%	%	5%	0%		
Howard County									
$r=0$	5	2	1	3	3	3	99		
	3%	6%	1%	%	%	%	%		
$0 < r \leq 10$	2	6	1	3	1	3	10		
	1%	0%	2%	%	%	%	0%		
$0 < r \leq 10$	1	3	3	1	4	6	10		
	6%	3%	0%	2%	%	%	1%		
$20 < r \leq 30$	1	1	1	2	1	1	99		
	3%	7%	7%	0%	3%	9%	%		
$30 < r \leq 40$	1	2	8	2	1	3	10		
	7%	%	%	1%	7%	5%	0%		
$r > 40$	9	2	2	7	4	7	10		
	%	%	%	%	%	6%	0%		
Total	2	3	1	6	4	1	10		
	8%	5%	3%	%	%	4%	0%		

Table 9. Percent of blocks in MSG and SF1, by percent renter and county (Continued)

SF1	MSG						To tal
	r= 0	0 <r≤10	1 0<r≤20	2 0<r≤30	3 0<r≤40	r> 40	
Queen Anne's County							
r=0	5 6%	1 1%	9 %	5 %	6 %	1 2%	99 %
0<r≤10	2 7%	4 2%	2 2%	3 %	2 %	4 %	10 0%
0<r≤10	2 6%	2 5%	2 1%	1 3%	5 %	1 0%	10 0%
20<r≤30	2 3%	1 1%	3 0%	1 5%	1 1%	1 0%	10 0%
30<r≤40	1 8%	2 %	2 0%	2 2%	2 0%	1 8%	10 0%
r>40	3 9%	2 %	1 0%	1 5%	8 %	2 6%	10 0%
Total	3 6%	1 8%	1 7%	1 0%	7 %	1 2%	10 0%
Hanover County							
r=0	5 9%	2 5%	9 %	3 %	1 %	3 %	10 0%
0<r≤10	3 0%	5 1%	1 3%	3 %	0 %	3 %	10 0%
0<r≤10	2 1%	4 3%	2 0%	9 %	0 %	6 %	99 %
20<r≤30	2 3%	2 7%	2 1%	1 0%	5 %	1 3%	99 %
30<r≤40	3 2%	6 %	2 6%	2 4%	3 %	9 %	10 0%
r>40	2 0%	7 %	8 %	5 %	1 4%	4 5%	99 %
Total	3	3	1	6	2	9	99

	5%	3%	4%	%	%	%	%
Henrico County							
$r=0$	4 9%	2 2%	1 4%	6 %	4 %	5 %	10 0%
$0 < r \leq 10$	2 2%	4 3%	2 1%	9 %	4 %	2 %	10 1%
$0 < r \leq 10$	1 6%	2 3%	3 0%	1 7%	9 %	4 %	99 %
$20 < r \leq 30$	1 1%	1 5%	2 7%	2 7%	1 2%	8 %	10 0%
$30 < r \leq 40$	7 %	5 %	2 4%	2 1%	2 4%	1 9%	10 0%
$r > 40$	8 %	1 %	4 %	5 %	7 %	7 5%	10 0%
Total	2 4%	2 3%	1 9%	1 1%	7 %	1 6%	10 0%

NOTES: Table may not add to 100 percent due to rounding.

The variable 'r' is the percentage of the occupied housing units in the block that are rental.

Table 10. Percent of blocks in Dunhill and SF1, by percent renter and county

SF1	Dunhill						Total
	$r=0$	$0 < r \leq 10$	$0 < r \leq 20$	$0 < r \leq 30$	$0 < r \leq 40$	$r > 40$	
Baltimore County							
$r=0$	4 5%	3 0%	1 6%	4 %	2 %	4 %	101 %
$0 < r \leq 10$	1 7%	6 0%	1 7%	4 %	1 %	2 %	101 %
$0 < r \leq 10$	1 6%	3 8%	3 2%	9 %	3 %	3 %	101 %
$20 < r \leq 30$	1 4%	2 2%	3 9%	1 5%	6 %	4 %	100 %

$30 < r \leq 40$	1 4%	1 2%	2 5%	2 5%	1 2%	1 1%	99%
$r > 40$	6 %	2 %	6 %	9 %	1 0%	6 7%	100 %
Total	2 2%	3 5%	1 9%	7 %	4 %	1 3%	100 %

Table 10. Percent of blocks in Dunhill and SF1, by percent renter and county (Continued)

SF1	Dunhill						Total
	r=0	0 <r≤10	1 0<r≤20	2 0<r≤30	3 0<r≤40	r>40	
Howard County							
r=0	4 6%	3 4%	1 2%	3 %	2 %	2 %	99%
0<r≤10	1 3%	6 7%	1 4%	2 %	1 %	2 %	99%
0<r≤10	1 8%	3 8%	3 0%	6 %	2 %	6 %	100%
20<r≤30	1 2%	1 4%	3 0%	2 0%	7 %	1 6%	99%
30<r≤40	1 9%	6 %	2 1%	1 9%	1 9%	1 7%	101%
r>40	1 1%	2 %	7 %	6 %	1 0%	6 4%	100%
Total	2 4%	4 0%	1 6%	5 %	4 %	1 1%	100%
Queen Anne's County							
r=0	6 1%	8 %	1 5%	6 %	4 %	5 %	99%
0<r≤10	2 3%	4 9%	1 9%	5 %	1 %	2 %	99%
0<r≤10	3 3%	2 5%	2 3%	8 %	6 %	5 %	100%
20<r≤30	4 4%	1 1%	2 5%	1 1%	4 %	4 %	99%
30<r≤40	4 1%	9 %	1 8%	1 6%	5 %	1 1%	100%
r>40	5 2%	2 %	1 0%	1 2%	4 %	2 1%	101%

Total	4	1	1	8	4	7	100
	4%	9%	8%	%	%	%	%
Hanover County							
r=0	5	2	1	3	2	3	101
	7%	4%	2%	%	%	%	%
0<r≤10	2	5	1	4	1	3	100
	4%	5%	3%	%	%	%	%
0<r≤10	2	3	2	8	2	2	100
	9%	9%	0%	%	%	%	%
20<r≤30	2	3	2	8	1	7	100
	6%	0%	8%	%	%	%	%
30<r≤40	3	6	3	6	3	1	98%
	5%	%	5%	%	%	3%	
r>40	2	4	1	1	1	2	100
	5%	%	9%	2%	4%	6%	%
Total	3	3	1	6	3	6	101
	5%	4%	7%	%	%	%	%
Henrico County							
r=0	5	2	1	3	1	2	100
	2%	8%	4%	%	%	%	%
0<r≤10	2	5	1	3	0	1	100
	7%	1%	8%	%	%	%	%
0<r≤10	2	3	2	7	1	1	100
	7%	7%	7%	%	%	%	%
20<r≤30	2	3	3	8	5	2	101
	2%	2%	2%	%	%	%	%
30<r≤40	2	1	3	1	5	8	100
	2%	9%	3%	3%	%	%	%
r>40	1	4	1	9	1	5	100
	2%	%	0%	%	1%	4%	%
Total	3	3	2	6	3	1	101
	1%	1%	0%	%	%	0%	%

NOTES: Table may not add to 100 percent due to rounding.

The variable 'r' is the percentage of the occupied housing units in the block that are rental.

The last set of tables of this nature are the distributions for the MSG and Dunhill files broken down by the size of the block in terms of the number of occupied housing units computed from the SF1 data. Table 11 shows the distribution for the MSG data. A relatively consistent pattern is that as the block size increases the percentages in the main diagonal increase, with only a few minor departures. In other words, the blocks with larger numbers of occupied units are more accurately classified with the MSG data. The pattern for the Dunhill data given in Table 12 is similar, with one exception. For blocks with no renters in the SF1, the percentages classified as having no renters by the Dunhill data decrease as the block size increases. It is possible that this may have something to do with the way the Dunhill data are classified by renter status, but we are uncertain about how this process works.

Table 11. Percent of blocks in MSG and SF1, by percent renter and block size

SF1		MSG					Total	
		$r=0$	$0 < r \leq 10$	$0 < r \leq 20$	$0 < r \leq 30$	$0 < r \leq 40$		$r > 40$
≤ 10	BLKSIZE=0< s							
	$r=0$	6 0%	8 %	1 3%	5 %	5 %	9 %	1 00%
	$0 < r \leq 10$	3 8%	2 0%	2 8%	4 %	4 %	6 %	1 00%
	$0 < r \leq 10$	3 9%	1 0%	1 9%	1 4%	7 %	1 0%	9 9%
	$20 < r \leq 30$	4 4%	1 1%	1 7%	1 1%	7 %	9 %	9 9%
	$30 < r \leq 40$	3 3%	2 %	2 6%	1 2%	1 0%	1 7%	1 00%
	$r > 40$	3 5%	3 %	8 %	8 %	8 %	3 9%	1 01%
Total		5 0%	8 %	1 5%	7 %	6 %	1 5%	1 01%
$s \leq 30$	BLKSIZE=10< s							
$r=0$	4	3	1	5	2	1	1	

	8%	2%	3%	%	%	%	01%
0<r≤10	3	3	2	7	3	2	1
	2%	5%	1%	%	%	%	00%
0<r≤10	2	2	2	1	6	4	1
	0%	8%	9%	3%	%	%	00%
20<r≤30	1	2	3	2	9	7	9
	1%	1%	0%	1%	%	%	9%
30<r≤40	6	1	2	2	1	1	1
	%	1%	2%	6%	7%	9%	01%
r>40	3	2	7	1	1	6	9
	%	%	%	1%	3%	3%	9%
Total	3	2	2	1	5	8	1
	0%	8%	0%	0%	%	%	01%
BLKSIZE=31							
or more							
r=0	3	5	1	2	1	1	1
	0%	6%	0%	%	%	%	00%
0<r≤10	1	6	1	5	2	2	1
	4%	1%	6%	%	%	%	00%
0<r≤10	4	4	2	1	5	3	1
	%	0%	9%	9%	%	%	00%
20<r≤30	3	1	2	3	1	1	1
	%	5%	5%	2%	5%	0%	00%
30<r≤40	3	5	1	1	3	3	1
	%	%	3%	8%	2%	0%	01%
r>40	1	1	2	2	3	9	1
	%	%	%	%	%	1%	00%
Total	1	4	1	9	4	2	1
	0%	0%	5%	%	%	2%	00%

NOTES: Table may not add to 100 percent due to rounding.

The variable 'r' is the percentage of the occupied housing units in the block that are rental. The variable 's' is the number of occupied housing units in the block.

Table 12. Percent of blocks in Dunhill and SF1, by percent renter and block size

SF1	Dunhill						Total
	r=0	0<r≤10	10<r≤20	20<r≤30	30<r≤40	r>40	
BLKSIZE=0<s≤10							
r=0	6 4%	8 %	1 4%	5 %	3 %	7 %	10 1%
0<r≤10	4 6%	1 7%	2 6%	6 %	3 %	3 %	10 1%
0<r≤10	5 2%	7 %	1 7%	1 3%	6 %	6 %	10 1%
20<r≤30	4 3%	9 %	2 3%	1 2%	8 %	7 %	10 2%
30<r≤40	4 2%	3 %	1 7%	1 3%	9 %	1 5%	99 %
r>40	3 8%	2 %	1 0%	8 %	8 %	3 3%	99 %
Total	5 4%	7 %	1 5%	8 %	5 %	1 2%	10 1%
BLKSIZE=10<s≤30							
r=0	4 4%	3 5%	1 7%	3 %	1 %	1 %	10 1%
0<r≤10	3 3%	4 1%	1 9%	5 %	1 %	1 %	10 0%
0<r≤10	2 5%	3 6%	2 6%	8 %	2 %	2 %	99 %
20<r≤30	2 1%	2 9%	3 1%	1 2%	4 %	3 %	10 0%
30<r≤40	1 5%	2 2%	3 4%	1 4%	9 %	7 %	10 1%
r>40	8 %	6 %	1 7%	1 5%	1 1%	4 4%	10 1%
Total	3	3	2	6	2	5	99

	1%	4%	1%	%	%	%	%
BLKSIZE=31 or more							
r=0	1 7%	7 4%	7 %	1 %	1 %	0 %	10 0%
0<r≤10	8 %	7 3%	1 5%	2 %	1 %	2 %	10 1%
0<r≤10	3 %	5 0%	3 7%	6 %	2 %	2 %	10 0%
20<r≤30	3 %	2 8%	4 7%	1 4%	4 %	5 %	10 1%
30<r≤40	1 %	1 1%	3 0%	3 6%	1 1%	1 1%	10 0%
r>40	1 %	2 %	3 %	7 %	1 1%	7 6%	10 0%
Total	6 %	4 9%	1 8%	6 %	4 %	1 8%	10 1%

NOTES: Table may not add to 100 percent due to rounding.

The variable 'r' is the percentage of the occupied housing units in the block that are rental. The variable 's' is the number of occupied housing units in the block.

In addition to these cross-tabulations, we computed some ratios at the block level to provide a more complete description of the accuracy of the lists. The three ratios computed for each list were: (1) the ratio of the number of occupied housing units in the list to the number of occupied housing units in the SF1; (2) the ratio of the number of rented units in the list to the number of rented units in the SF1; and, (3) the ratio of the number of owned units in the list to the number of owned units in the SF1. One important note concerning these tables is the summaries of the ratios given below are done at the block-level. Thus, each block is treated equally, irrespective of the number of housing units in the block.

Table 13 gives the ratios of the number of occupied housing units from the lists to the SF1 overall and by state and county. The mean of the ratios for the MSG and the Dunhill are both greater than unity, indicating overcoverage. This finding is consistent with the results in Table 1.⁷ As noted earlier, the main source of variation in the ratios for both the MSG and Dunhill files are the counties rather than aggregates by state.

⁷ The means are not identical to ratios that could be computed from Table 1 because each block is counted equally in the means.

More interesting than the means are the distributions of the ratios for the two lists. The medians from the MSG file are always close to unity, while those for the Dunhill file are always less than unity. Looking at all the percentiles, the distribution for the MSG ratios are basically shifted upward compared to the Dunhill, often by about 0.10 to 0.25. The standard deviation is also an interesting measure of the stability of the ratios. The standard deviation of the ratio for Dunhill is larger than for MSG, but this appears to be largely a function of the data for Henrico county in Virginia. In this county, the Dunhill file distribution is severely shifted downward (the median ratio is 0.47) and yet there are some large ratios (the 90th percentile ratio is 2.0). The standard deviation for this county is 12.6 and this causes the Dunhill ratios to be less stable for the aggregates at the Virginia and overall level.

Table 14 gives the ratios of the number of rented housing units from the lists to the SF1 overall and by state and county. Table 15 gives the corresponding ratios of the number of owned units. The ratios for the rented units in Table 14 are not as stable as those for the owned units in Table 15 because the denominator (the number of rented units in the SF1 file) may be small and this will cause the ratio to be unstable. Nonetheless, the pattern in Table 14 for rented units is consistent with that observed in Table 13 for all occupied units. The Dunhill file has severe problems for Henrico county. However, this pattern does not persist for the ratios of the number of owned units given in Table 15. These ratios are relatively stable and the Dunhill ratios are closer to the MSG ratios. In particular, the ratios for Henrico county from the Dunhill file are very similar to the MSG ratios and are not out of line with the ratios for other counties. Thus, the discrepancy in the Dunhill file is due to the differences in both the number of all occupied and the number of rented units in Henrico county, not the number of owned units in the county.

Table 13. Ratios of the number of occupied housing units in a block for MSG and Dunhill to SFI, by geography

Geography	List	Mean	Median	Percentiles				
				50	55	60	65	70
All	MS	.31	.79	.82	.20	.00	.84	.56
	Dunhill	.17	.55	.55	.03	.88	.69	.47
MD	MS	.33	.18	.83	.20	.00	.84	.56
	Dunhill	.16	.74	.60	.05	.88	.70	.49
VA	MS	.26	.70	.75	.20	.00	.85	.57
	Dunhill	.19	.93	.50	.00	.86	.67	.46
Baltimore County	MS	.26	.22	.67	.17	.00	.86	.63
	Dunhill	.11	.21	.46	.00	.88	.72	.50
Howard County	MS	.60	.17	.22	.36	.04	.86	.56
	Dunhill	.36	.68	.90	.19	.92	.74	.50
Queen Anne's County	MS	.43	.74	.33	.42	.93	.50	.26
	Dunhill	.31	.42	.00	.26	.80	.50	.25
Hanover County	MS	.49	.64	.50	.50	.00	.71	.40
	Dunhill	.21	.78	.00	.25	.87	.60	.36
Henrico County	MS	.18	.28	.58	.14	.00	.88	.64

	Dunhill	.19	0.15	.29	.00	.85	.70	.49	0
--	---------	-----	------	-----	-----	-----	-----	-----	---

Table 14. Ratios of the number of rented housing units in a block for MSG and Dunhill, by geography

Geography	List	Mean	Median	Percentiles						
				50	55	60	65	70	75	80
All	MSG	.57	.92	.00	.50	.92	.00	.00	.00	
	Dunhill	.48	.61	.50	.00	.60	.00	.00	.00	
MD	MSG	.60	.36	.00	.57	.92	.00	.00	.00	
	Dunhill	.58	.90	.00	.22	.67	.09	.00	.00	
VA	MSG	.50	.85	.00	.50	.93	.00	.00	.00	
	Dunhill	.26	1.40	.00	.00	.48	.00	.00	.00	
Baltimore County	MSG	.62	.79	.00	.58	.93	.04	.00	.00	
	Dunhill	.59	.39	.00	.20	.67	.20	.00	.00	
Howard County	MSG	.86	.00	.00	.94	.00	.00	.00	.00	
	Dunhill	.78	.34	.27	.67	.91	.17	.00	.00	
Queen Anne's County	MSG	.03	.85	.00	.00	.50	.00	.00	.00	
	Dunhill	.17	.53	.00	.00	.25	.00	.00	.00	
Hanover County	MSG	.15	.79	.08	.00	.50	.00	.00	.00	
	Dunhill	.27	.69	.00	.00	.50	.00	.00	.00	

Henrico County	G	MS	.61	.32	.00	.67	.00	.17	.00	.00
		Dunhill	.25	2.63	.00	.00	.47	.00	.00	

Table 15. Ratios of the number of owned housing units in a block for MSG and Dunhill, by geography

Geography	List	Mean	Median	Percentiles						
				10	25	50	75	90	95	100
All	MS	.50	.85	.00	.26	.00	.82	.52		
	Dunhill	.54	.25	.00	.16	.92	.75	.50		
MD	MS	.53	.80	.14	.27	.00	.83	.53		
	Dunhill	.59	.43	.00	.16	.92	.75	.50		
VA	MS	.43	.95	.00	.25	.00	.81	.52		
	Dunhill	.43	.81	.00	.17	.92	.75	.50		
Baltimore County	MS	.54	.14	.00	.23	.00	.85	.60		
	Dunhill	.65	.71	.00	.14	.92	.77	.55		
Howard County	MS	.55	.53	.50	.40	.03	.84	.57		
	Dunhill	.45	.53	.19	.22	.94	.76	.50		
Queen Anne's County	MS	.42	.44	.50	.47	.00	.50	.24		
	Dunhill	.29	.02	.20	.30	.87	.50	.25		
Hanover County	MS	.63	.98	.67	.52	.00	.73	.37		

County	Henrico	Dunhill	.29	.57	.00	.27	.90	.66	.36
		MS	.37	.94	.79	.20	.00	.83	.57
		Dunhill	.47	.88	.93	.13	.93	.77	.57

Overall, the analysis suggests that the MSG data correspond more closely to the SF1 data than the Dunhill data when the data are aggregated to the block-level. The variability of the data with respect to the SF1 data by county raises some concerns about whether the findings can be generalized to other states and counties not included in this analysis.

3.3 Assessment Using CPI Data

The other source of data that can be used to evaluate the data from the lists is the block-level summary data from the CPI Housing Survey as provided by BLS. As we noted earlier, the main limitation associated with using these data is that they are sample data and do not cover many of the blocks in the areas and even those included are only covered for a sample of housing units.

We attempt to deal with the small sample size in two ways. First, instead of examining the more complete distribution of the percent of renters as done in the previous section, we classify each block into either a 40 percent or less renter category or a more than 40 percent renter category. Second, we create three groups of blocks depending on the number of sampled housing units that were found in the CPI sample. The three groups are less than 5 sampled occupied units, 5 to 9 sampled occupied units, and 10 or more sampled occupied units. Any categorization of a block based on a small sample size is obviously tenuous so the small and medium categories are not very informative for this analysis. Unfortunately, most of the blocks fall into these less informative categories. There are 147 blocks with less than 5 sampled units, 72 with between 5 and 9 sampled units, and only 61 with 10 or more sampled units.

An even more troublesome problem for the analysis is the fact that there are very few rental units in the blocks with more than 10 sampled units in the CPI. Table 16 shows the number of blocks cross-classified by the percent renter in the CPI and the SF1 by block size. This table only counts the number of blocks in the CPI and the SF1 that match. The appendix contains tables that include the five blocks that did not match when merged with the SF1 file.

Table 16. Number of blocks in CPI and SF1, by percent renter and CPI sample size

SF1	CPI		T otal
	0 <r≤40	r> 40	
All blocks			
0<r	2	7	2
≤40	00		07
r>4	2	5	7
0	1	2	3
ss<5			
0<r	8	6	9
≤40	7		3
r>4	1	3	5
0	6	8	4
5≤ss			
0<r	5	1	5
≤40	5		6
r>4	2	1	1
0		4	6
ss>9			
0<r	5	0	5
≤40	8		8
r>4	3	0	3
0			

NOTES: Table may not add to 100 percent due to rounding.

The variable 'r' is the percentage of the occupied housing units in the block that are rental.

The variable 'ss' is the number of housing units in the CPI sample.

Because of the sample size limitations, we restrict our analysis to overall aggregates by the CPI block sample size. The appendix has more complete tables of the counts by state, county, and block size and the appendix tables include the missing data for those interested in the details of the matching of the CPI and SF1 files. Table 17 shows the percent distribution of the blocks by the percent renter for the CPI and the SF1. As with the earlier data, when coarse categories are used, the percentage agreement is high.

Tables 18 and 19 have the same format as Table 17, but instead of the SF1 data the MSG and Dunhill data are tabulated. The percentages in the main diagonal cells are high for these tables, but not quite as large as they are in Table 17 using the SF1 file data. The main difference is that both lists are less accurate for blocks that are identified as having more than 40 percent renters in the CPI. However, this difference cannot be given much credence if we assume the SF1 data are more accurate than the CPI data.

Table 17. Percent of blocks in CPI and SF1, by percent renter and CPI sample size

SF 1	CPI		T otal
	0 <r≤40	r> 40	
All blocks			
0 <r≤40	9 6.6%	3 4%	1 00.0%
r >40	2 9.5%	7 0.5%	1 00.0%
ss< 5			
0 <r≤40	9 3.5%	6 5%	1 00.0%
r >40	3 1.0%	6 9.0%	1 00.0%
5≤ ss<10			

	0	9	1.	1
< <i>r</i> ≤40	8.2%		8%	00.0%
	<i>r</i>	1	8	1
>40	1.8%	8.2%		00.0%
	<i>ss</i> >			
9				
	0	1	0	1
< <i>r</i> ≤40	00.0%		%	00.0%
	<i>r</i>	1	0	1
>40	00.0%	%		00.0%

NOTES: Table may not add to 100 percent due to rounding.

The variable '*r*' is the percentage of the occupied housing units in the block that are rental.

The variable '*ss*' is the number of housing units in the CPI sample.

Table 18. Percent of blocks in CPI and MSG, by percent renter and CPI sample size

MS G	CPI		T otal
	0 <r≤40	r> 40	
All blocks			
0 <r≤40	9 4.4%	5. 6%	1 00.0%
r >40	4 2.2%	5 7.8%	1 00.0%
5 ss<			
0 <r≤40	9 0.0%	1 0.0%	1 00.0%
r >40	3 8.6%	6 1.4%	1 00.0%
5≤ ss<10			
0 <r≤40	9 6.0%	4. 0%	1 00.0%
r >40	4 0.9%	5 9.1%	1 00.0%
9 ss>			
0 <r≤40	1 00.0%	0 %	1 00.0%
r >40	1 00.0%	0 %	1 00.0%

NOTES: Table may not add to 100 percent due to rounding.

The variable 'r' is the percentage of the occupied housing units in the block that are rental.

The variable 'ss' is the number of housing units in the CPI sample.

Table 19. Percent of blocks in CPI and Dunhill, by percent renter and CPI sample size

MS G	CPI		T otal
	0 <r≤40	r> 40	
All blocks			
0	9	8.	1
<r≤40	1.5%	5%	00.0%
r	3	6	1
>40	9.7%	0.3%	00.0%
5 ss<			
0	8	1	1
<r≤40	8.7%	1.3%	00.0%
r	3	6	1
>40	4.0%	6.0%	00.0%
5≤ ss<10			
0	8	1	1
<r≤40	7.9%	2.1%	00.0%
r	4	5	1
>40	2.9%	7.1%	00.0%
9 ss>			
0	1	0	1
<r≤40	00.0%	%	00.0%
r	1	0	1
>40	00.0%	%	00.0%

NOTES: Table may not add to 100 percent due to rounding.

The variable 'r' is the percentage of the occupied housing units in the block that are rental.

The variable 'ss' is the number of housing units in the CPI sample.

4. Summary

Lists of housing units by census block are available from vendors and these lists might be used to enhance or replace an in-person listing and screening process currently employed in the CPI Housing Survey. To evaluate the quality of the lists, we purchased lists from MSG and Dunhill for five counties in the Baltimore and Richmond Metropolitan Statistical Areas. The costs of obtaining the lists were \$15,000 for the MSG list and \$11,000 for the Dunhill list.

This report compares the data from the lists with the corresponding block-level data from the 2000 Decennial Census as contained in the SF1 data file. Since the CPI Housing Survey is especially interested in blocks with 40 percent or less renters, the comparisons focused on the distributions of the percent renters. In addition, BLS provided us with block-level data from the CPI Housing Survey in the five counties and these data are also summarized in this report.

The MSG file contained 557,559 housing unit records from the five specified counties and 63 percent of these units have a telephone number available. The Dunhill file contained 471,868 records from the five specified counties and 58% of these housing units have a telephone number available.

The first issue considered in the report was the coverage of the lists. After matching the lists to the SF1 file, the MSG list had either about the same or more housing units than the SF1 file for each of the five counties, while list from Dunhill had fewer housing units than the MSG list and the SF1 file in each county. These comparisons suggest that the gross coverage rate for the MSG file is better than that of the Dunhill file, but the Dunhill coverage rate is still relatively good.

The second evaluation was of the accuracy of the percent renter classifications from the two vendor files as compared to the SF1 data. The Dunhill list differed by more than 6 percentage points from the SF1 percent in three of the five counties, while the MSG list never differed from the SF1 percent by more than 2 percentage points when aggregates were examined. The analysis was then restricted to the matching blocks and again the MSG list appeared to be more accurate. The largest differences between the MSG and Dunhill percent renter distributions were for the Census blocks with 0 to 10 percent renter, where the list from Dunhill was more accurate, and for categories with more than 20 percent renters where the MSG list was more accurate.

The percent renter distributions by state and county were also examined. While the distributions were very similar by state, the county distributions were rather odd. Two counties had very

low match rates as compared to other counties. This finding could indicate a local component to be the quality of the lists.

Further analysis of the ratios of the number of housing units and rented units for the two lists also produced an unusual finding at the county level. The list from Dunhill for Henrico county in Virginia did not track with the SF1 data. This difference caused some of the summaries of aggregates at higher levels such as the state level to be poorer for the Dunhill list.

The analysis of the percent renter distribution indicated the MSG data correspond more closely to the SF1 data than the Dunhill data when the data are aggregated to the block-level. However, the variability of the data by county raises some concerns about whether the findings can be generalized to other states and counties not included in this analysis.

The last analysis used the CPI Housing Survey data. The small sample size severely limited the ability to use these data to evaluate the vendor lists. We attempted to deal with the small sample size by using coarser percentage renter distributions and aggregating only to groups based on the sample size in the CPI blocks. Despite these efforts, no definitive conclusions can be drawn from these data with respect to the quality of the lists of the vendors.

The next task in the project involves selecting a sample of household from the lists and comparing the data from the list for the housing unit to data collected from the sampled housing units. An important goal of this task is to examine the accuracy of tenure data from the lists at the housing unit level. The report from that task will greatly supplement the information from this analysis.

APPENDIX A

DETAILED TABLES

Table A-1. Number of blocks in MSG and SFI when codes 0 to 5 are treated as renters, by percent renter

S	MSG							T
	r=0	0<r≤10	0<r≤20	0<r≤30	0<r≤40	r>40	M	
F1							issing	otal
r=0	1,724	825	425	152	97	155	495	3,873
0<r≤10	783	1,724	650	215	79	67	44	3,562
0<r≤10	370	644	582	332	127	93	66	2,214
20<r≤30	148	163	244	215	103	84	44	1,001
30<r≤40	80	34	114	104	98	113	26	569
r>40	189	34	76	90	119	1,277	258	2,043
T	3,294	3,424	2,091	1,108	623	1,789	933	13,262

NOTES: The variable 'r' is the percentage of the occupied housing units in the block that are rental.

Table A-2. Number of blocks in MSG and SFI when codes 0 to 6 are treated as renters, as percent renter

S	MSG							To
	r=0	0<r≤10	0<r≤20	0<r≤30	0<r≤40	r>40	Mi	
F1							issing	tal
r=0	1,307	838	657	206	151	219	495	3,873
0<r≤10	400	1,500	1,021	363	145	89	44	3,562
0<r≤10	217	391	727	461	207	145	66	2,214
20<r≤30	95	71	218	262	179	132	44	1,001

	30	4	2	8	1	1	1	26	56
< <i>r</i> ≤40		7	0	5	10	26	55		9
	<i>r</i> >	1	1	6	7	1	1,	25	2,
40		38	9	6	8	10	374	8	043
total	T	2,	2,	2,	1,	9	2,	93	13
		204	839	774	480	18	114	3	,262

NOTES: The variable '*r*' is the percentage of the occupied housing units in the block that are rental.

Table A-3. Number of blocks in MSG and SFI, by state and percent renter

SF1	MSG							Total
	r=0	0<r≤10	10<r≤20	20<r≤30	30<r≤40	r>40	Missing	
MD								
r=0	1,158	574	282	98	63	109	347	2,631
0<r≤10	545	1,274	464	146	53	43	32	2,557
0<r≤10	243	435	373	217	78	57	50	1,453
20<r≤30	97	98	145	125	63	49	30	607
30<r≤40	57	23	64	61	57	77	22	361
r>40	132	21	52	61	75	893	172	1,406
Total	2,232	2,425	1,380	708	389	1,228	653	9,015
VA								
r=0	566	251	143	54	34	46	148	1,242
0<r≤10	238	450	186	69	26	24	12	1,005
0<r≤10	127	209	209	115	49	36	66	761
20<r≤30	51	65	99	90	40	35	14	394
30<r≤40	23	11	50	43	41	36	4	208
r>40	57	13	24	29	44	384	86	637
Total	1,062	999	711	400	234	561	280	4,247

NOTES: The variable 'r' is the percentage of the occupied housing units in the block that are rental.

Percent number based on codes 0 to 5.

Table A-4. Number of blocks in MSG and SFI, by county and percent renter

SF1	MSG							Total
	r=0	0<r≤10	10<r≤20	20<r≤30	30<r≤40	r>40	Missing	
Baltimore County								
r=0	815	445	217	72	36	67	201	1,853
0<r≤10	411	944	376	128	44	23	17	1,943
0<r≤10	175	341	290	176	63	31	27	1,103
20<r≤30	693	77	108	99	45	28	13	439
30<r≤40	40	21	50	40	39	51	9	250
r>40	72	16	37	33	59	756	97	1,070
Total	1,582	1,844	1,078	548	286	956	364	6,658
Howard County								
r=0	206	101	42	13	12	12	60	446
0<r≤10	93	266	54	14	6	14	10	457
0<r≤10	25	52	48	19	6	10	12	172
20<r≤30	9	12	12	14	9	13	6	75
30<r≤40	8	1	4	10	8	17	4	52
r>40	12	3	3	10	6	105	32	171
Total	353	435	163	80	47	171	124	1,373
Queen Anne's County								
r=0	137	28	23	13	15	30	86	332
0<r≤10	41	64	34	4	3	6	5	111

0									57
0	0<r≤1	43	42	35	22	9	16	1	1
0	20<r≤	19	9	25	12	9	8	1	78
30	30<r≤	9	1	10	11	10	9	1	9
40	r>40	48	2	12	18	10	32	9	3
								4	5
								3	9
									1
									65
	Total	29	14	13	80	56	10	1	9
		7	6	9			1	65	84
	Hanover County								
	r=0	15	66	25	8	3	8	9	3
		9						6	65
0	0<r≤1	82	13	36	7		7	7	2
			8						77
0	0<r≤1	38	76	36	16		11	9	1
									86
30	20<r≤	18	21	16	8	4	10	8	8
									5
40	30<r≤	11	2	9	8	1	3	1	3
									5
	r>40	20	7	8	5	14	45	4	1
								2	41
	Total	32	31	13	52	22	84	1	1
		8	0	0				63	,089
	Henrico County								
	r=0	40	18	11	46	31	38	5	8
		7	5	8				2	77
0	0<r≤1	15	31	15	62	26	17	5	7
		6	2	0					28
0	0<r≤1	89	13	17	99	49	25	7	5
			3	3					75
30	20<r≤	33	44	83	82	36	25	6	3
									09
40	30<r≤	12	9	41	35	40	33	3	1
									73
	r>40	37	6	16	24	30	33	4	4
							9	4	96
	Total	73	68	58	34	21	47	1	3
		4	9	1	8	2	7	17	,158

NOTES: The variable 'r' is the percentage of the occupied housing units in the block that are rental.

Percent number based on codes 0 to 5.

Table A-5. Number of blocks in MSG and SFI, by block size and percent renter

SF1	MSG							Total
	r=0	0<r≤10	10<r≤20	20<r≤30	30<r≤40	r>40	Missing	
BLKSIZE=0								
<s≤10								
r=0	8	1	1	6	6	1	4	1
	78	21	93	7	8	30	38	,895
0<r≤10	2	1	2	3	3	4	7	7
	7	4	0					8
0<r≤10	1	3	7	5	2	3	4	4
	46	8	1	0	7	8	0	10
20<r≤30	9	2	3	2	1	1	3	2
	2	3	5	3	5	9	0	37
30<r≤40	6	4	4	2	1	3	2	2
	2		9	3	9	2	2	11
r>40	1	1	3	3	4	1	1	6
	68	3	6	6	0	85	71	49
Total	1	2	4	2	1	4	7	3
	,373	13	04	02	72	08	08	,480
BLKSIZE=10								
<s≤30								
r=0	7	4	1	7	2	2	5	1
	28	83	91	7	7	0	0	,576
0<r≤10	4	5	3	1	4	3	3	1
	88	24	15	07	5	2	3	,544
0<r≤10	1	2	2	1	5	3	2	9
	87	67	71	25	5	3	3	61
20<r≤30	4	9	1	9	4	3	1	4
	8	4	33	3	1	3	2	54
30<r≤40	1	2	4	5	3	4	4	2
	4	4	8	7	7	2		26
r>40	1	8	2	3	4	2	3	3
	2		4	7	6	19	4	80
Total	1	1	9	4	2	3	1	5

	,477	,400	82	96	51	79	56	,141
BLKSIZE=30 or more								
r=0	18	21	41	81	22	55	77	402
0<r≤10	68	,186	15	05	11	31	44	,940
0<r≤10	73	39	40	57	55	22	33	843
20<r≤30	88	46	76	99	47	32	22	310
30<r≤40	44	66	17	24	42	39	33	132
r>40	93	13	16	17	33	88	53	1014
Total	444	,811	057	104	002	,002	96	,641

NOTES: The variable 'r' is the percentage of the occupied housing units in the block that are rental. The variable 's' is the number of occupied housing units in the block.

Percent number based on codes 0 to 5.

Table A-6. Number of blocks in Dunhill and SFI, by percent renter

SF	Dunhill							Total
	r=0	0<r≤10	10<r≤20	20<r≤30	30<r≤40	r>40	Missing	
0	1,632	925	479	117	54	114	486	3,873
r≤10	680	2,027	589	117	25	58	6	3,562
0<r≤10	452	779	617	169	54	58	50	2,214
20<r≤30	185	230	324	118	47	41	37	1,001
30<r≤40	108	69	145	101	50	57	32	569
r>40	213	47	137	156	179	1,023	185	2,043

tal	To	3,	4,	2,	77	40	1,	79	13,
		270	077	291	8	9	351	6	262

NOTES: The variable 'r' is the percentage of the occupied housing units in the block that are rental.

Table A-7. Number of blocks in Dunhill and SFI, by state and percent renter

SF1	Dunhill							Total
	r=0	0<r≤10	10<r≤20	20<r≤30	30<r≤40	r>40	Missing	
MD								
r=0	1,053	632	335	83	42	88	398	2,631
0<r≤10	421	1,514	427	88	19	45	43	2,557
10<r≤20	248	503	428	117	43	88	66	1,453
20<r≤30	100	112	206	77	22	11	93	607
30<r≤40	60	36	79	77	40	40	29	361
r>40	135	26	75	106	117	755	192	1,406
Tot	2,017	2,823	1,550	558	293	1,007	767	9,015
VA								
r=0	579	293	144	34	12	26	154	1,242
0<r≤10	259	513	162	99	63	13	33	1,005
10<r≤20	204	276	189	52	11	0	9	761
20<r≤30	85	118	118	31	55	0	7	394
30<r≤40	48	33	66	24	10	17	0	208

40	$r >$	7 8	2 1	6 2	5 0	6 2	2 68	9 6	6 37
al	Tot	1, 253	1, 254	7 41	2 20	1 16	3 44	3 19	4, 247

NOTES: The variable 'r' is the percentage of the occupied housing units in the block that are rental.

Table A-8. Number of blocks in Dunhill and SFI, by county and percent renter

SF1	Dunhill							Total
	r=0	0<r≤10	10<r≤20	20<r≤30	30<r≤40	r>40	Missing	
Baltimore County								
r=0	743	487	255	58	26	67	217	1,853
0<r≤10	329	1,146	335	99	11	22	12	1,943
10<r≤20	168	404	344	66	11	0	0	1,103
20<r≤30	57	93	165	44	24	7	19	439
30<r≤40	33	92	16	11	9	7	0	50
r>40	62	21	54	85	100	646	102	1,070
Total	1,392	2,180	1,214	333	221	819	399	6,658
Howard County								
r=0	173	126	46	22	7	9	7	446
0<r≤10	58	294	33	11	6	0	5	574
10<r≤20	28	60	48	9	3	1	1	172
20<r≤30	88	10	21	14	5	1	6	177
30<r≤40	9	30	10	9	9	8	4	155
r>40	15	30	10	80	13	86	36	371
Total	291	496	198	63	43	134	48	1,373
Queen Anne's County								
r=0	1	1	3	1	9	1	1	3

	37	9	4	3		2	08	32	
0<r≤10	3	7	2	8	2	3	7	1	
	4	4	9					57	
0<r≤10	5	3	3	1	9	8	2	1	
	2	9	6	2			2	78	
20<r≤30	3	9	2	9	3	3	1	9	
	5		0				4	3	
30<r≤40	1	4	8	7	2	5	1	5	
	8						5	9	
r>40	5	2	1	1	4	2	5	1	
	8		1	3		3	4	65	
Total	3	1	1	6	2	5	2	9	
	34	47	38	2	9	4	20	84	
County									
Hanover									
r=0	1	6	3	7	5	8	9	3	
	54	4	2				5	65	
0<r≤10	6	1	3	1	3	9	8	2	
	5	47	4	1				77	
0<r≤10	5	6	3	1	3	4	1	1	
	0	9	5	4			1	86	
20<r≤30	1	2	2	6	1	5	1	8	
	9	2	1				1	5	
30<r≤40	1	2	1	2	1	4	4	3	
	1		1					5	
r>40	2	4	1	1	1	2	4	1	
	4		8	1	3	5	6	41	
Total	3	3	1	5	2	5	1	1	
	23	08	51	1	6	5	75	,089	
County									
Henrico									
r=0	4	2	1	2	7	1	5	8	
	25	29	12	7		8	9	77	
0<r≤10	1	3	1	1	3	4	1	7	
	94	66	28	8			5	28	
0<r≤10	1	2	1	3	8	6	8	5	
	54	07	54	8				75	
20<r≤30	6	9	9	2	1	5	6	3	
	6	6	7	5	4			09	
30<r≤40	3	3	5	2	9	1	6	1	
	7	1	5	2		3		73	

r>40	5	1	4	3	4	2	5	4
	4	7	4	9	9	43	0	96
Total	9	9	5	1	9	2	1	3
	30	46	90	69	0	89	44	,158

NOTES: The variable 'r' is the percentage of the occupied housing units in the block that are rental.

Table A-9. Number of blocks in Dunhill and SFI, by block size and percent renter

SF1	Dunhill							Total
	r=0	0<r≤10	0<r≤20	0<r≤30	0<r≤40	r>40	Missing	
BLKSIZE=0 <s≤10								
r=0	901	106	192	73	41	96	486	1,895
0<r≤10	33	12	19	44	22	22	65	78
0<r≤10	186	24	26	64	12	11	05	410
20<r≤30	86	18	45	23	15	13	37	237
30<r≤40	75	55	31	24	17	27	32	211
r>40	178	10	48	38	38	152	185	649
Total	1,459	175	397	208	134	311	796	3,480
BLKSIZE=10 <s≤30								
r=0	666	527	260	40	10	17	56	1,576
0<r≤10	498	615	288	73	11	9	04	1,544
0<r≤10	237	339	244	73	8	0	03	961
20<r≤30	91	128	135	53	9	4	14	454
30<r≤40	32	49	75	30	9	5	6	226
r>40	27	19	58	49	36	148	43	380

Total	1,551	1,677	1,060	318	113	233	189	5,141
BLKSIZE=30 or more								
r=0	65	292	27	44	32	11	10	402
0<r≤10	149	1,400	82	0	2	7	0	1,940
10<r≤20	29	416	311	0	5	7	5	843
20<r≤30	88	84	144	2	3	4	5	310
30<r≤40	11	15	93	7	4	5	1	132
r>40	88	18	31	96	05	23	60	1,014
Total	260	2,225	348	52	62	07	01	4,641

NOTES: The variable 'r' is the percentage of the occupied housing units in the block that are rental. The variable 's' is the number of occupied housing units in the block.

Table A-10. Number of blocks in CPI and SF1, by percent renter

	CPI 0<r≤40			CPI r>40			Total
	SF1 r			SF1 r			
	0<r≤40	r>40	Missing	0<r≤40	r>40	Missing	
Total	200	21	2	7	52	3	285
MD	104	14	1	1	30	2	52
VA	96	7	1	6	22	1	133
ss=1	87	16	2	6	38	2	151
ss=2	55	2		1	14	1	73

Total	1 94	2 7	2	1 8	4 1	3	2 85
MD	1 01	1 7	1	6	2 5	2	1 52
VA	9 3	1 0	1	1 2	1 6	1	1 33
ss=1	8 6	1 7	2	1 1	3 3	2	1 51
ss=2	5 1	6		7	8	1	7 3
ss=3	5 7	4					6 1

NOTES: The variable 'r' is the percentage of the occupied housing units in the block that are rental.

The variable 'ss' is the number of housing units in the CPI sample.

**EVALUATION OF VENDOR LISTS:
COMPARISON OF VENDOR AND
INTERVIEW DATA**

FINAL REPORT

Submitted to:

Bureau of Labor Statistics
2 Massachusetts Ave, NE
Washington D.C 20212

Submitted by:

Westat
1650 Research Blvd.
Rockville, MD 20850

February 26, 2003



TABLE OF CONTENTS

<u>Section</u>		<u>Page</u>
1	INTRODUCTION.....	1
2	METHODS.....	1
	2.1 Sample Design.....	2
	2.2 Questionnaire.....	5
	2.3 Interviewer Training.....	5
	2.4 Data Collection.....	5
	2.5 Weighting and Estimation of Variances.....	6
	2.6 Interviewing Results.....	10
3	RESULTS.....	11
	3.1 Quality of Contact Information.....	11
	3.2 Quality of Tenure Status.....	12
4	DISCUSSION.....	17

List of Appendices

		<u>Page</u>
A	SUPPLEMENTARY TABLES.....	A-1
B	TELEPHONE QUESTIONNAIRE.....	B-1
C	IN-PERSON QUESTIONNAIRE.....	C-1
D	TRAINING AGENDA.....	D-1
E	ADVANCE LETTER.....	E-1

List of Tables

<u>Table</u>		<u>Page</u>
1	Universe Totals for Sample Frame.....	19
2	Final Result by Mode of Interview.....	20
3	Final Result by Match Between Vendors.....	21
4a	Final Result by MSG Tenure Status.....	22
4b	Final Result by Collapsed MSG Tenure Status.....	23
5	Final Result by Dunhill Tenure Status.....	24
6	Final Result by Percent Renters on Block.....	25
7	MSG Tenure Status by Survey Report of Tenure Status.....	26

TABLE OF CONTENTS (CONTINUED)

List of Tables (Continued)

<u>Table</u>		<u>Page</u>
8	MSG Tenure Status by Survey Report of Tenure Status for the Two Sample Groups.....	27
9	Collapsed MSG and Dunhill Tenure Status by Survey Report of Tenure Status for the Two Sample Groups.....	28
10	MSG Tenure Status by Survey Report of Tenure Status for Mode Groups.....	30
11	Collapsed MSG and Dunhill Tenure Status by Survey Report of Tenure Status for Mode Groups.....	31
12	MSG Tenure Status by Survey Report of Tenure Status for Percent Renters on Block.....	33
13	Collapsed MSG and Dunhill Tenure Status by Survey Report of Tenure Status for Percent Renters on Block.....	35
14	MSG Tenure Status by Survey Report of Tenure Status for Each County.....	39
15	Collapsed MSG and Dunhill Tenure Status by Survey Report of Tenure Status for Each County.....	42

1. Introduction

The Office of Prices and Living Conditions of the Bureau of Labor Statistics (BLS) is exploring the use of purchased lists of addresses to enhance or replace the in-person listing and screening processes used to identify renters in the Consumer Price Index (CPI) Housing Survey. If a list can be used to identify renters, it is anticipated that there could be some cost savings associated with the CPI survey. Perhaps more importantly, it may provide a way to enhance the survey's ability to identify renters in high owner areas --- a difficult problem for the CPI in the past.

To conduct the research, we purchased two lists of housing units for selected counties in the Baltimore and Richmond Metropolitan Statistical Areas. The lists were purchased from Marketing Systems Group (MSG) and Dunhill International. The MSG list cost \$15,000 and the Dunhill list was \$11,000. Both the MSG and Dunhill lists use U.S. Postal delivery addresses as the base and append additional data from other sources. The prime data source for the MSG list is Info USA and Dunhill uses Knowledgebase.

A previous report provided information on the accuracy of these two lists at a block level (Westat, 2002). The purpose of this report is to describe an evaluation of each list at an individual housing unit level. The next section provides an overview of the methods used to collect the data. The third section describes the results of the evaluation and the final section summarizes the results.

2. Methods

This project focussed on evaluating the quality of information on tenure status provided by the MSG and Dunhill lists (hereafter referred to as "tenure status"). This variable is being considered for use sampling for the CPI Housing survey. There is interest to use the variable to improve the efficiency of identifying renters in high-owner neighborhoods. To conduct the evaluation, a sample of housing units was drawn in selected Maryland and Virginia counties. These states were chosen because they were relatively close to Westat, which allowed for some cost savings when conducting the interviews. In Baltimore, the counties were Baltimore County, Howard County and Queen Anne's County. In Virginia, the counties were Hanover County and Henrico County.

The sample was selected by dividing the households into those with and those without a telephone number, as indicated by the vendor for each unit. Samples were drawn from each of these groups with the intent of having field interviewers make attempts to contact and collect information on the ownership status of the household. The field interviewers were instructed to conduct a telephone interview for those units that had a telephone number and conduct an in-person interview with those where no telephone number was provided.

A short, 10-item questionnaire was administered to each household. The primary focus of the interview was to assess tenure status. These data were then used to assess the quality of the information provided by the vendor files.

2.1 Sample Design

The first step in sampling for the project was to create a sampling frame for selecting households. The sampling frame was created from the data files from the vendors (MSG and Dunhill) for the two selected MSAs, but was further restricted as described below. The purpose of limiting the coverage of the frame further in the following way: (1) only blocks where the Census reported between 0 and 40% renters were included (blocks with 0 renters were excluded); (2) only households in the MSG file were included (if a household was only identified from the Dunhill file it was excluded); (3) housing units in the MSG file with a tenure value from '0' to '8' were included (households with a tenure value of '9' on the MSG file, is most likely to be an owned unit, were excluded); (4) housing units with no telephone number listed were included only those in block groups with a large enough number of housing units. Also households with no telephone in and Queen Anne's county in Maryland were excluded.

These restrictions were imposed to efficiently estimate the reliability of the vendor files in blocks that had a high proportion of owners. In addition, these restrictions kept the costs of the data collection to an acceptable level.

The initial step of frame construction was matching the data files from Dunhill and MSG at the housing unit level. This was done using matching software. The matching specifications required both units to be in the same census block (thus, state, county, tract, and block had to be identical). Matching then took place using address, name and telephone number

from the two vendor files. Priority in the statistical matching was given to units whose address in both files matched exactly. Records with similar addresses and the same telephone or name were also considered a match.

From this matching, three files were created corresponding to records that appeared on both MSG and Dunhill (M&D), MSG only (M-o), and Dunhill only (D-o). As noted above, units from the D-o file were eliminated from the sampling frame, because the project team was unsure of the quality of the information on the Dunhill file at the time of sampling. Table 1 provides the number of households in each of the groups (M&D; M-o; D-o). The M&D and the M-o files were concatenated at this point.

The next step was to delete housing units from the frame when the MSG variable indicating ownership status had the highest possible value of '9' (indicating a high degree of certainty that unit was owned). The reason for this restriction was that previous analysis by BLS found that the '9' value was highly reliable, in the sense that it matched closely with actual data from the CPI survey. The second row in Table 1 provides the universe totals after taking out records where the MSG tenure variable had a value of '9'. Eliminating these units substantially changes the composition of the universe. Over half (65%) of the records are eliminated from the universe. The cases deleted are disproportionately in the group where the two vendor files match one another.

The next step was to stratify the housing units by the presence of a telephone number (although we sometimes refer to these as telephone and nontelephone strata, the stratification is based on whether the vendor file contained a telephone number for the household rather than whether the household actually had a telephone). The third and fourth rows of Table 1 give the universe counts in the two strata. The last step in creating the sampling frame was to exclude from the nontelephone strata those housing units that were either in Queen Anne's county or were in block groups with 11 or fewer households. The next to last row in Table 1 provides the count for the restricted nontelephone stratum and the last row gives the final universe totals overall (across both telephone and nontelephone strata).

The telephone stratum was further partitioned by county and tenure status (owned or rented). If the MSG tenure value was less than 6 the household was classified as a renter and if it was greater than 5 it was classified as an owner. Essentially, the telephone stratum was divided

into 10 strata (the five counties crossed by the two tenure status). The sample was then allocated to each stratum with the goal of obtaining 600 completed telephone interviews after accounting for response rates and eligibility rates. The sample size allocated was the same for each tenure status in a county (e.g., in Baltimore 243 cases were sampled from the owned stratum and 243 cases were sampled from the rented stratum). The total sample size selected by county was: 486 in Baltimore, 202 in Howard, 250 in Queen Anne's, 142 in Hanover, and 420 in Henrico. Within the 10 strata the households were selected with simple random sampling. Originally, a subsample of 500 of the total 1,500 sampled cases were selected to be a reserve sample. Later, due to lower than anticipated yields, the entire telephone reserve sample was released.

The design of the nontelephone sample differed from the simple stratified design of the telephone stratum in order to reduce data collection costs. A two-stage sample design was used, where 16 block groups were selected from Maryland (across Baltimore and Howard counties) and 10 block groups were selected from Virginia (Hanover and Henrico counties). The block groups were sampled within each state using probabilities proportional to the number of households in the block group using systematic sampling after sorting the block groups by county.

In the second stage of sampling, households were sampled from the sampled block groups. As in the telephone sample, households were first classified as owned or rented and an equal size sample of 8 owned households and 8 rented households were selected from each block group by simple random sampling. If the block group had fewer than 8 owned (or rented) households, then all of the units in that tenure were included in the sample. Overall, 401 households were sampled (15 fewer than would have been selected if each block group had enough units of the appropriate tenure). The final step was to include about one-third of the cases into the reserve sample. To do this, the sample cases were sorted by state, block group, and tenure status and an equal probability, systematic sample was selected. Of the 401 cases only 268 were released to the field and the remaining 133 were reserve cases that were never released.

As noted above, the initial sample released for interviewing was 1,268 households, 1,000 from the telephone stratum and 268 without a telephone number. Midway through the field period the lower yield rates required releasing the additional 500 cases with telephone numbers to achieve the targeted number of 800 completed interviews. Thus, the final sample size was 1,768 households.

2.2 Questionnaire

Based on questions proposed by BLS, Westat developed two questionnaires – one for the telephone cases (Appendix B) and one for the in-person cases (Appendix C). They differ in that the telephone version verifies the respondent's address, and also asks if the address is for a business. After identifying a respondent who lives in the home and is at least 18 years old, the survey asks whether the home is rented or owned, and how much the rent or mortgage is.

2.3 Interviewer Training

Seven field interviewers and one field supervisor were trained. Interviewers received 10 hours of training – four home study and 6 in-person. In-person training was held October 14, 2002 at Westat's Rockville offices. Roughly one week before that date, interviewers were sent a general interviewer training manual that describes Westat interviewing procedures. Interviewers were asked to read this manual and complete test questions about the material. The in-person session consisted of lecture and role plays. Topics included an overview of the project and purpose of the survey; administering the questionnaires; contact procedures, and administrative procedures. The training agenda appears in Appendix D. The field supervisor attended the same training session as the field interviewers, but also met with project staff separately to discuss supervisory tasks and responsibilities.

2.4 Data Collection

Data collection began October 15, 2002. The field period was originally scheduled to end November 26, 2002. However, data collection was extended to December 31, 2002 in order to ensure that the second sample release received sufficient contact attempts.

The field supervisor held weekly one-on-one telephone conversations with each interviewer to review outstanding cases, discuss any unusual situations the interviewer may have encountered, and advise the interviewer of any procedural updates. In turn, the field supervisor

participated in weekly conference calls with Westat project staff to report case status and any outstanding interviewer issues.

All sampled persons were sent an advance letter (Appendix E) on BLS letterhead. The letter notified them that the study would be taking place, and that they would be contacted by an interviewer for a very brief interview. All sampled persons who completed an interview were sent a thank you letter and brief self-administered survey, the purpose of which was to validate that they had completed the interview as reported by the field interviewer.

About halfway through data collection, it became evident that the field interviewers were having difficulty with the telephone cases. Therefore, on November 11, 2002, the Westat Telephone Research Center (TRC) began calling the telephone cases which field interviewers had been unable to contact. In addition, the reserve sample consisted of telephone cases only, and those were assigned directly to the TRC. Calls to reserve sample households began on November 25, 2002.

For the field cases, we sent a short survey to all households that completed an interview, asking them to verify that a field interviewer had contacted them to ask questions about their ownership status and mortgage or rental amount. For a sample of 10 percent of the telephone cases, the field supervisor placed a follow up call to the household if the survey was not returned.

2.5 Weighting and Estimation of Variances

The tables in this report contain estimates from the survey. The estimates are based on the weighted counts of the number of households with different characteristics, where the weights account for sampling from the frame and contain an adjustment for nonresponse. The first step of weighting was to produce a base weight that is the inverse of the probability of selecting the household from the frame. Note that these base weights do not contain any adjustments for households that were eliminated from the sampling frame as described above.

In the 10 telephone strata (defined above as the cross of county and tenure status), the households were selected by simple random sampling. Thus, the inverse of the probability of selection for every household within a county by tenure status is equal to the number of units in the sampling frame in that stratum divided by the number sampled. The weight can be written as

$$w_{t,hi} = \frac{N_{t,hi}}{n_{t,hi}}$$

where $N_{t,hi}$ is the number of households in the sampling frame in telephone stratum in county h and tenure status i , $n_{t,hi}$ is the number of sampled households in telephone stratum in county h and tenure status i .

In the nontelephone stratum, the baseweight is the product of the two stages of sampling. The probability of selecting a block group in a state is proportional to $S_j/S.$, where S_j is the number of households in block group j and $S.$ is the sum of the S_j over all the nontelephone households in the sampling frame. At the second stage, the probability of selecting a household in tenure status k in sampled block group j is $8/T_{jk}$ where T_{jk} is the number of households in block group j that are in tenure status k . Thus, the baseweight is the product of the inverse of these two terms and can be written as

$$w_{nt,jk} = \frac{S.}{pS_j} \frac{T_{jk}}{8}$$

where p is the number of block groups selected in the state. Since the reserve sample was not released for field work, the final baseweight is the number of households released in tenure status k rather than the fixed number 8 presented above.

If responses had been obtained for all sampled households, estimates using the baseweights would be unbiased. However, a nonresponse adjustment to the baseweights was used to compensate because some households did not respond. All the sampled and released households were divided into three categories (respondents, nonrespondents and ineligible) based on their participation in the survey. These three categories are described below.

Category 1: **Respondents.** This group consists of all eligible sample units that participated in the survey.

Category 2: **Nonrespondents.** This group consists of all eligible sample units that did not provide substantially complete and usable survey data.

Category 3: **Ineligibles.** This group consists of all sample units that were ineligible or out of scope for the survey.

To reduce the bias of the estimates, the nonresponse adjustments were computed for households within adjustment classes or cells that were relatively homogeneous with respect to response rates. The cells were formed by examining the response rates for several characteristics that were available from the sampling frame. The table below defines the non-response adjustment cells.

Cell	Mode	State	MSG Tenure Value	A ^(nr)
1	Telephone number	D	M 0,1,2,3	6 1.7018
2	Telephone number	D	M 4,5,6	5 2.2143
3	Telephone number	D	M 7,8	7 1.6399
4	Telephone number	A	V 0,1,2,3	0 1.3769
5	Telephone number	A	V 4,5,6	2 1.5630
6	Telephone number	A	V 7,8	3 1.4539
7	No Telephone number	D	M ALL VALUES	1 1.2048
8	No Telephone number	A	V ALL VALUES	3 1.1372

Within each of the 8 cells, the baseweight was multiplied by a nonresponse adjustment factor. The nonresponse adjustment factor is the ratio of the sum of the baseweights for respondents and nonrespondents to the sum of the baseweights for the respondents. The nonresponse adjustment factor for cell c can be written as

$$A_c^{(nr)} = \frac{\sum_{i \in R_c} w_i + \sum_{i \in N_c} w_i}{\sum_{i \in R_c} w_i}$$

where w_i is baseweight for household i (omitting the subscripts), R_c is the set of responding households in cell c , and N_c is the set of nonresponding households in cell c . The factors are given in the table above. The nonresponse adjusted weight for household i (w'_i) is

$$\begin{aligned} w'_i &= A_c^{(nr)} w_i && \text{if } i \in c \text{ and is a responding household,} \\ w'_i &= 0 && \text{otherwise.} \end{aligned}$$

These nonresponse adjusted weights are used in the report.

In order to compute the precision of the estimates, weights that can be used with replication variance software were also created. A total of 126 replicate weights using the following procedures. A total of 12 variance strata were created, 10 were for the telephone strata (defined by county and tenure status) and 2 were for the nontelephone strata (defined by state). Within each of the 10 telephone strata, the sampled households were sorted in the order of selection and systematically assigned to variance units labelled 1 to 10 (thus there were 10 variance units in each of the 10 variance strata for the telephone strata. In the nontelephone strata, the households in the same primary sampling unit (the block groups) were assigned to the same variance unit (thus there were 16 variance units in Maryland and 10 in Virginia). Consequently, every sampled household was assigned to one of 12 variance strata and one of either 10 variance units (for the telephone strata) or 26 variance units (for the nontelephone households).

Using this structure, the replicate weights were created using a stratified jackknife procedure in a standard fashion. Within a variance stratum the replicate baseweight was created by deleting the base weight for all households in the same variance unit (by making it zero) and

increasing the baseweight for the other households in the same variance stratum. The weights for the other strata are not altered. Since there were 10 variance strata with 10 variance units in the telephone strata, this results in 100 replicate weights. The remaining 26 replicate weights are associated with the 16 and 10 variance units in the nontelephone strata. For example, consider the first replicate weight which is associated with the first telephone strata and variance unit. Replicate weight 1 for all households in this first telephone strata are set to zero if they are for variance unit 1. Replicate weight 1 for households from the same telephone stratum but a different variance unit are the household's baseweight times 10/9 (the number of variance units in the stratum divided by one less than this number). Replicate weight 1 for all other households that are not in the first variance stratum are equal to their baseweight. The process for the other replicate weights follows the same procedure.

Since the weights were adjusted for nonresponse, the same nonresponse adjustment method was used to create replicate nonresponse adjusted weights. Essentially, the nonresponse adjustment factors $A_c^{(nr)}$ were first computed for each of the 126 replicate weights and then these adjustments were multiplied by the corresponding base replicate weights to produce the nonresponse adjusted replicate weights.

The precision of the estimates were then be computed using these replicate weights. We used WesVar (version 4) and the JK_n method (this corresponds to the stratified replication method described above). The variance of an estimate using this replication method can be written as

$$v_{jkn} = \sum_{h=1}^H \frac{n_h - 1}{n_h} \sum_{i=1}^{n_h} \left(\hat{\theta}_{(hi)} - \hat{\theta}_h \right)^2$$

where $\hat{\theta}_{(hi)}$ is the replicate estimate for stratum h when variance unit i is deleted (computed using replicate nonresponse adjusted weight i) and $\hat{\theta}_h$ is the estimate from stratum h using the full sample nonresponse adjusted weight.

2.6 Interviewing Results

Table 2 shows final result codes for the total sample of 1768 and by mode of interview. If one includes “wrong or nonexistent address” as a complete, then the overall response rate is 69.7%, with 86.1% in-person and 66.2% on the telephone. These rates exclude businesses and non-working numbers from the denominator. If one excludes the bad addresses from both the numerator and denominator, the overall response rate is 66.6%, with 59.2% on the telephone and 84.9% for the in-person component.

There is some variation in response rates across different types of units. For example, the response rate varied by whether the two vendor files matched on the address (Table 3), with those units that matched having a higher response rate than those that didn’t match (66.6% vs. 58.3%). Table 4a shows the response rates by the full set of MSG tenure codes. Table 4b collapses the tenure codes using “0-2” to be renters and “3-8” as owners. When doing this, owners have a higher response rate than renters (64.8% vs. 55.5%). Interestingly, the opposite seems to be the case when using the Dunhill tenure indicator (Table 5), although these data only apply to units that matched with the MSG file. There is no discernable pattern in the response rates by the percent of renters on the block (Table 6).

3. Results

Two different types of analyses were conducted. The first examined the quality of the contact information provided on the files. The second looked at the quality of the tenure information. The latter was assessed by comparing the tenure status provided by the two different vendor files to what was collected when completing the interview.

3.1 Quality of Contact Information

Both the address and the telephone number for the sampled unit were provided in the files. To evaluate the address information, the results listed in Table 2 were used. For the in-person interviews, these data were used to calculate the proportion of housing units where there was an incorrect address. There were a total of 268 in-person sample units. In all cases, the interviewer was in a position to either confirm or reject the address provided by the MSG file.

Using these data, it is estimated that approximately 8.2 percent of the addresses in the file do not exist. This takes the 22 cases where a wrong address was identified and divides it by 268.

The telephone portion of the sample provides information on the quality of the telephone numbers provided in the files. One measure of quality is the number of telephone numbers that are for a business. In addition to the code for a business, the interview results used in this calculation are the non-working numbers and completed interviews. None of the other result codes could be used to determine whether the unit called was a business or not. From this, it is estimated that approximately 4 percent were for a business. A second measure of quality is the proportion of non-working numbers. There were a total of 1500 numbers in this portion of the sample. Of these, 186 were non-working or 12.4 percent.

Finally, interviewers asked each telephone respondent whether the address provided by the vendor was the same as the address of the respondent. For 220 cases, the respondent noted the address was different. The denominator for this proportion includes all those cases where it could be determined that the address was correct, including the businesses and the completed interviews. Using these as a base, approximately 25 percent of the telephone numbers were for the wrong address.

3.2 Quality of Tenure Status

Tables 7 through 15 provide a comparison of the information that was obtained during the interview with that indicated on each of the two vendor files. Each of these tables presents the percent of cases comparing the two measures. The percentages were calculated with data that were weighted as described above in section 2.5. Standard errors and significance tests were estimated using the procedure described above. The standard error for the percentages were suppressed if the denominator of the percentage was based on less than 20 unweighted cases.

Many of the comparisons are in the form of a 2 x 2 tables comparing the tenure status of the data file and the survey report. For these tables, two different statistics are presented. One is the Gross Difference Rate (GDR). This is a measure of disagreement between the vendor and interview. It is interpreted as the total number of households that were

misclassified by the vendor record (assuming the interview is correct). For each 2x2 table of the counts:

		Inte	
		O	R
endor	V wn	a	b
	R ent	c	d

the GDR is computed as:

$$\text{GDR} = (b+c)/n. \text{ Where } n=a+b+c+d$$

The 2x2 tables presented below are expressed as weighted percentages of the total. Also included in the tables are the standard errors of this percentage, the weighted and unweighted counts. One can compute the GDR from these percentages as the sum of the two off-diagonal elements (b+c).

The second measure listed in these table is the Net Difference Rate (NDR). This is a measure of the net bias in the vendor data . It is computed by:

$$\text{NDR}=(b-c)/n.$$

This can be computed from the tables discussed below by subtracting the off-diagonal elements, again using the weighted percentages.

Table 7 compares the full MSG tenure status variable with the survey report. The MSG data have a total of 9 values, ranging from 0 to 8. The low value (0) represent units that the source is most confident as being a renter, while the high value is where the source is most confident to be an owner.⁸

As can be seen from Table 7, the number of cases for the low values of the MSG tenure variable are quite small. This is partly because the blocks were selected from the sample

⁸ The original MSG file had a value of '9'. As noted in the sample design section, these cases were eliminated from the sample frame.

that had a relatively high percentage of owners. The relationship between the survey report and the MSG tenure status is statistically significant (chi-square = 26, df=8; $p < .0001$). As the MSG values get higher, there is a greater chance that the survey report will be an “owner”. The match with the survey report seems to have a point of inflection where MSG is equal to ‘3’. In this case, 73.1% reported on the survey to be owning. This compares to values of less than 40% for the 0-2 values.

Table 8 provides these data broken out by the matching status between the two vendor files. The pattern evident on Table 7 holds for the sample where the MSG and the Dunhill file match. The pattern is slightly different for the sample where there are no matches. For this portion of the sample, there seem to be many more owners, with only one value having a majority of self-reported renters (tenure status = 4).

Table 9 provides the comparison for each of the vendor files. To make this comparison, the MSG tenure variable was collapsed into two categories. Renters were defined as those with values 0-2, while owners were defined with values 3-8. These values were selected because of the pattern noted in Table 7 above, which found a point of inflection in self-reported ownership between the “2” and “3” value. Table 9 breaks out the comparison by whether or not there was a match between the two vendor files. Where there was a match, the MSG has a GDR value of 21.9, indicating that these data did not match the survey report 21.9% of the time. A large portion of the disagreement was when the MSG indicated the household was being owned, but the interview file indicated it was actually rented. The latter bias is quantified by a large positive NDR of 16.9%. That is, the error rate was higher for vendor identified owner units than renter units. This NDR is significantly different from zero, as indicated by the relatively small standard error (2.7).

A similar pattern emerges when doing this comparison for the MSG data that did not match to the Dunhill file. These data are less likely to correspond to the survey reports at all, with a GDR of 37.2%. This is significantly different from the group that did match (difference in GDR's = 15.3; $p < .05$). The NDR for the group that did not match (30.6%) is in the same direction for the group that did match and is quite a bit larger (difference in NDR's = 13.7; $p < .05$).

A different pattern emerges for the Dunhill file. Of the households identified as owners by this source, 28.5% disagreed on tenure status, with the largest error being when Dunhill indicated a renter, but the interview listed the house as being owned. In other words, the NDR for the Dunhill file is in a different direction than for the comparable MSG data. This is further reinforced when comparing the percent of units identified as an owner by the vendor that were consistent with the survey information. The percent of units identified by the MSG as a renter is 71.9% ($6.4/8.9 = 71.9$), compared to 45.7% for the Dunhill file ($14.5/31.7 = 45.7$). As shown at the bottom of the table, this difference of 26.2% ($71.9 - 45.7 = 26.2$) is statistically significant ($z=3.4$).

More generally, the pattern of the GDR's across the vendor data-sets indicates that the MSG data are slightly more accurate than the Dunhill data ($p<.05$). The opposite is true for the NDRs ($p<.05$).

The difference in the direction of the NDR's is partly a function of the way the MSG tenure status variable was created. If one splits the MSG data by defining an "owner" using a higher value of this variable, then the NDR becomes negative (like Dunhill). When doing this, however, the GDR also goes up. For example, using the values of "0-5" to define a "renter" on the MSG file increases to the GDR to approximately 40% and shifts the NDR to -17%.

The comparison between the two data sources is complicated by the way the sample was selected. As noted in section 2.5, the units the MSG file identified as "certainty" owners (MSG code = 9) were eliminated from the frame. Taking these out of the current analysis, then, should lead to an overstatement of the total error in the files. In particular, it may drive the MSG file to a higher level of error, since the sampling was done after the most accurate records on the file had been eliminated.

The match between the vendor and interview data does vary slightly by the mode of the interview (Table 10). There is a consistent tendency for the in-person households to have a slightly higher percentage of owners across all groups. However, only one of these differences is statistically significant at conventional levels (value of "3" – 86.6 vs. 59.7; $p<.05$).

Comparison between the vendor files for the different modes is shown in Table 11, which restricts the sample to only those cases where the MSG and Dunhill files matched each other.⁹ For the Dunhill data, the GDR's are 31.6 and 25.4 for the in-person and telephone respectively. Similar, but slightly lower, GDR's occur for the MSG file (21.3 and 22.5). Neither of the differences in the GDRs between vendors is statistically significant at conventional levels, although the in-person difference approaches significance ($z=1.7$; $p<.10$). For the Dunhill telephone data, there appears to be very little or no net bias. The owners identified by Dunhill are about as equally likely to be mis-identified as the renters, as indicated by the NDR of $-.3$. It is not clear whether this is an effect of mode of interview or sample design. The households that had a telephone number in the vendor files were assigned to the telephone mode, while those without a telephone number were assigned to in-person interviewing. Mode and type of unit, therefore, are confounded in this comparison.

As with the total sample, the MSG file does a better job of identifying renters than the Dunhill file for both modes of interviews. The difference in the percent of renters identified by the MSG and Dunhill that have a survey indicating the unit is rented is 23.6% and 22.0% for the in-person and telephone, respectively (bottom of each panel of Table 11). These are significant at $p<.10$ ($z=1.7, 1.8$).

For the MSG data, the GDR are directly correlated with the percent of persons renting on the block (Table 13). For the MSG data, the GDRs range from 14.6 to 42.9 as one moves from low renter (0% - 10%) to high renter (31% - 40%). This pattern is not evident for the Dunhill file. The range of GDRs is from 25.8 (0% - 10% renter) to 29.0 (31% - 40% renter).

For the NDR's, the MSG data also show a direct relationship with the percent of renters on the block, although it is not quite as strong as found for the GDRs. The NDRs range from a low of 11.8 in the 0% - 10% blocks to 49.8 in the 31% - 40% blocks. The Dunhill data do not show as strong a pattern in this direction. The highest NDR is for the block with the most renters (19.6), but this has a relatively high standard error. This estimate is not statistically different from 0, using a 95% confidence interval.

⁹ See Appendix A for these tables for the comparisons with the MSG data for the entire sample.

The MSG data identify renters better than the Dunhill data for the low-renter blocks (0% - 10% renters). The percentage of renters identified by each vendor that is consistent with the survey 23.6% higher for the MSG data, which is statistically significant ($z=2.2$). The direction of this difference is the same for the other types of blocks, but only approaches statistical significance for the blocks with the most renters (31% - 40% renters; difference = 12.7; $z=1.6$).

The final set of tables are for the individual counties. The data disaggregated by the full MSG tenure variable have very small sample sizes, which makes it difficult to interpret (Table 14). For the collapsed MSG variable (Table 15), there is no clear pattern across the counties. For the GDRs, for example, no one county stands out as being particularly accurate or inaccurate. While there is more variation for the NDRs, these have larger standard errors and many are not significantly different from 0.

4. Discussion

How one views the quality of the information from the two vendor files depends on how they will be used. Assuming an in-person contact for all addresses, the data described above indicate that approximately 8 percent of the addresses will not exist. For the MSG file, between 21 percent and 37 percent (depending on the matching status to the Dunhill file) of the records indicating tenure status are in error, as judged by the tenure status reported during the interview. The overall bias depends on how the MSG tenure-status variable is collapsed. The approach taken above was to use a relatively low value of the MSG tenure status variable to serve as a cutoff for designating a unit as an owner. Under this scenario, the bias tended to be greatest among those that the vendor identified as an owner, but turned out to be a renter. However, if one uses a larger value of the MSG tenure variable to define an “owner” (e.g., “5”), then the bias is in the opposite direction (i.e., larger for units identified by the vendor as “renters”). When switching coding schemes in this way, the gross difference rate goes up as well (e.g., from around 25 to 40).

The Dunhill file exhibited higher gross difference rates and lower net difference rates than the MSG file. The direction of the net difference rates were also different across the two files, at least as defined in the above tables. The NDR for the MSG file tended indicate the error was largest for units identified as owners, while for the Dunhill file the error tended to be

greatest for units identified as renters. If one plans on using the file to identify renters, therefore, the MSG file is preferable.

For the MSG file, there was a correlation between the percent of renters on the block and the accuracy of the information. The more renters on the block, the greater the error. In addition, the direction of this error seemed to vary, with the “owner” designation of the MSG file being in error more often as the proportion of renters on the block increased. This pattern was not evident for the Dunhill file. Neither the gross or net error rates varied systematically by the percent of renters on the block. The MSG file seemed to best at identifying renters, when compared to the Dunhill file, for those blocks that had the fewest renters. If the greatest interest is to identify renters on those blocks where there are mostly owners, then the MSG file seems to perform the best relative to the Dunhill file.

Table 1. Universe Totals for Sample Frame.

	MSG & Dunhill Match	MSG only	Dunhi ll only	Total
<i>TOTAL</i>	266,795	63,083	29,898	359,776
<i>Total without MSG Tenure = 9</i>	65,435	30,045		95,480
Telephone number	35,953	15,101		51,054
No telephone number	29,482	14,944		44,426
No telephone number without Queen Anne's County	28,300	14,393		42,693
<i>Final Sample Frame</i>	64,253	29,494		93,747

Table 2. Final Result by Mode of Interview.

Result Code	Total Sample	Tele phone	In Person
Telephone Interview Completed	629	629	NA
In-person Interview Completed	208	NA	208
Business	33	32	1
Final Breakoff	2	2	0
Final Other	18	3	15
Final Refusal	203	190	13
Language Barrier	6	6	0
Maximum Telephone Call Attempts (12) Reached	80	80	NA
No Answer	156	150	6
No Eligible Respondent Found	5	2	3
Non-working Telephone Number	186	186	NA
Wrong Address/No Such Address	242	220	22
Total	1768	1500	268
<i>High response rate</i> ⁺	69.7 %	66.2 %	86.1 %
<i>Low response rate</i> *	64.0 %	59.2 %	84.9 %

⁺ = (Completes + Wrong Address) / (Final Breakoff + Final Other + Final Refusal + Language Barrier + Maximum Calls + No Answers + No Eligible Respondent + Completes + Wrong Address)

* (Completes) / (Final Breakoff + Final Other + Final Refusal + Language Barrier + Maximum Calls + No Answers + No Eligible Respondent + Completes)

Table 3. Final Result by Match Between Vendor Files.

Result Code	Address Match Between Vendors	No Address Match Between Vendors
Telephone Interview Completed	454	175
In-person Interview Completed	148	60
Business	18	15
Final Breakoff	1	1
Final Other	12	6
Final Refusal	145	58
Language Barrier	3	3
Maximum Telephone Call Attempts (12) Reached	49	31
No Answer	90	66
No Eligible Respondent Found	2	3
Non-working Telephone Number	103	83
Wrong Address/No Such Address	140	102
Total	1165	603
<i>High response rate</i> ⁺	71.1%	66.7%
<i>Low response rate</i> *	66.6%	58.3%

⁺ = (Completes + Wrong Address) / (Final Breakoff + Final Other + Final Refusal + Language Barrier + Maximum Calls + No Answers + No Eligible Respondent + Completes + Wrong Address)

* (Completes) / (Final Breakoff + Final Other + Final Refusal + Language Barrier + Maximum Calls + No Answers + No Eligible Respondent + Completes)

Table 4a. Final Result by MSG Tenure Status.

Result Code	Rent								O	
	0	1	2	3	4	5	6	7	8	
Telephone Interview Completed	8	6	9	2	7	6	9	2	30	70
In-person Interview Completed	0	4	9	6	1	3	0	2	4	8
Business	0	1	1	4	8	7	4	6	6	2
Final Breakoff	0	0	0	0	0	1	0	1	1	0
Final Other	0	0	2	1	2	7	2	3	3	1
Final Refusal	2	4	1	1	2	3	2	4	4	4
Language Barrier	0	2	0	0	0	2	1	0	0	1
Maximum Telephone Call Attempts (12) Reached	0	1	5	7	3	1	5	6	1	7
No Answer	1	2	3	1	1	2	2	2	3	2
No Eligible Respondent Found	1	0	1	0	0	0	0	2	0	1
Non-working Telephone Number	3	1	5	1	3	7	7	2	3	8

Wrong Address/ No Such Address	2	9	2	3	3	4	3	3	2
		6	2	8	4	2	0	9	
Total	17	30	11	11	22	22	22	33	33
	7	0	12	92	64	69	06	39	39
High response rate⁺	7	6	6	7	6	6	6	6	7
	1.4%	7.9%	6.7%	9.0%	8.5%	5.1%	9.2%	9.6%	0.5%
Low response rate*	6	5	5	7	6	5	6	6	6
	6.7%	2.6%	4.3%	3.6%	1.9%	7.1%	2.7%	6.2%	7.6%

⁺ = (Completes + Wrong Address) / (Final Breakoff + Final Other + Final Refusal + Language Barrier + Maximum Calls + No Answers + No Eligible Respondent + Completes + Wrong Address)

* (Completes) / (Final Breakoff + Final Other + Final Refusal + Language Barrier + Maximum Calls + No Answers + No Eligible Respondent + Completes)

Table 4b. Final Result by Collapsed MSG Tenure Status.

Result Code	Own	Rent
Telephone Interview Completed	586	43
In-person Interview Completed	195	13
Business	31	2
Final Breakoff	2	0
Final Other	16	2
Final Refusal	186	17
Language Barrier	4	2
Maximum Telephone Call Attempts (12) Reached	74	6
No Answer	140	16
No Eligible Respondent Found	3	2
Non-working Telephone Number	167	19
Wrong Address/ No Such Address	205	37
Total	1609	159
<i>High response rate</i> ⁺	69.9%	67.4%
<i>Low response rate</i> *	64.8%	55.5%

⁺ = (Completes + Wrong Address) / (Final Breakoff + Final Other + Final Refusal + Language Barrier + Maximum Calls + No Answers + No Eligible Respondent + Completes + Wrong Address)

* (Completes) / (Final Breakoff + Final Other + Final Refusal + Language Barrier + Maximum Calls + No Answers + No Eligible Respondent + Completes)

Table 5. Final Result by Dunhill Tenure Status.

Result Code	Own	Rent
Telephone Interview Completed	337	117
In-person Interview Completed	97	51
Business	9	9
Final Breakoff	1	0
Final Other	8	4
Final Refusal	121	24
Language Barrier	2	1
Maximum Telephone Call Attempts (12) Reached	34	15
No Answer	64	26
No Eligible Respondent Found	2	0
Non-working Telephone Number	65	38
Wrong Address/No Such Address	86	54
Total	826	339
<i>High response rate</i> ⁺	69.2%	76.0%
<i>Low response rate</i> *	65.2%	70.6%

⁺ = (Completes + Wrong Address) / (Final Breakoff + Final Other + Final Refusal + Language Barrier + Maximum Calls + No Answers + No Eligible Respondent + Completes + Wrong Address)

* (Completes) / (Final Breakoff + Final Other + Final Refusal + Language Barrier + Maximum Calls + No Answers + No Eligible Respondent + Completes)

Table 6. Final Result by Percent Renters on Block.

Result Code	0-10%	11-20%	21-30%	31-40%
Telephone Interview Completed	291	180	107	51
In-person Interview Completed	114	59	19	16
Business	12	15	5	1
Final Breakoff	1	1	0	0
Final Other	9	8	1	0
Final Refusal	101	62	23	17
Language Barrier	2	2	0	2
Maximum Telephone Call Attempts (12) Reached	35	23	14	8
No Answer	69	51	17	19
No Eligible Respondent Found	3	1	1	0
Non-working Telephone Number	70	54	33	29
Wrong Address/No Such Address	98	61	50	33
Total	805	517	270	176
High response rate⁺	69.6%	67.0%	75.9%	68.5%
Low response rate[*]	64.8%	61.8%	69.2%	59.3%

⁺ = (Completes + Wrong Address) / (Final Breakoff + Final Other + Final Refusal + Language Barrier + Maximum Calls + No Answers + No Eligible Respondent + Completes + Wrong Address)

^{*} (Completes) / (Final Breakoff + Final Other + Final Refusal + Language Barrier + Maximum Calls + No Answers + No Eligible Respondent + Completes)

Table 7. MSG Tenure Status by Survey Report of Tenure Status.

	0	1	2	3	4	5	6	7	8	Total*
Own	0.1%	4.9%	7.4%	3.1%	8.7%	1.85	2.0%	8.9%	3.3%	1.7%
Standard error	-	-	(9.7)	(6.4)	(6.3)	(6.2)	(4.4)	(4.5)	(3.4)	(2.1)
Unweighted n	2	2	12	61	74	77	75	133	163	599
Weighted N	85	380	1506	6013	7296	6714	7784	10460	11889	52126
Rent	9.9%	5.1%	2.6%	6.9%	1.3%	8.2%	8.0%	1.1%	6.8%	8.3%
Standard error	-	-	(9.7)	(6.4)	(6.3)	(6.2)	(4.4)	(4.5)	(3.4)	(2.1)
Unweighted n	6	8	25	29	33	27	19	43	27	217
Weighted N	337	709	2519	2209	3323	2641	1711	4724	2392	0564
Total*	00%	00%	00%	00%	00%	00%	00%	00%	00%	00%
Unweighted n	8	10	37	90	107	104	94	176	190	816
Weighted N	442	1088	4025	8223	10618	9355	9495	15184	14281	72690

* Totals may not add up due to rounding.

- Denominator of percent is less than 20 unweighted cases.

Table 8. MSG Tenure Status by Survey Report of Tenure Status for the Two Sample Groups (Column Percents).

	Rent						Own					
	0	1	2	3	4	5	6	7	8	9	Total*	
Address Match Between Vendors												
Own												
Standard error	24.4%	9.4%	2.4%	7.7%	6.2%	6.1%	6.3%	9.3%	6.5%	4.2%		
Unweighted n	-	-	(10.7)	(6.5)	(4.0)	(8.6)	(4.3)	(5.0)	(3.3)	(2.3)	(
Weighted N	2	1	6	36	50	54	54	99	134	436		
	85	73	1150	3745	5400	4987	5866	7957	9780	39043		
Rent												
Standard error	75.7%	0.6%	7.6%	2.3%	3.8%	3.9%	3.7%	0.7%	3.5%	5.8%		
Unweighted n	-	-	(10.7)	(6.5)	(4.0)	(8.6)	(4.3)	(5.0)	(3.3)	(2.3)	(
Weighted N	5	8	23	20	13	17	11	34	19	150		
	264	709	2398	1789	864	1566	929	3526	1529	13573		
Total*												
Unweighted n	100.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	
Weighted N	7	9	29	56	63	71	65	133	153	586		
	349	782	3548	5533	6264	6554	6795	11482	11309	52616		
No Address Match Between Vendors												
Own												
Standard error	0%	00.0%	4.5%	4.4%	3.5%	1.6%	1.1%	7.6%	1.0%	5.2%		
Unweighted n	-	-	-	(8.9)	(12.0)	(11.9)	(9.6)	(8.3)	(10.7)	(4.9)	(
Weighted N	0	1	6	25	24	23	21	34	29	163		
	0	306	356									

				2269	1895	1726	1918	2503	2109	13083
Rent	1	0	2	1	5	3	2	3	2	3
Standard error	00.0%	%	5.5%	5.6%	6.5%	8.4%	9.0%	2.4%	9.0%	4.8%
Unweighted n	-	-	-	(((((((
Weighted N	1	0	2	9	20	10	8	9	8	67
	73	0	122	421	2459	1075	782	1198	863	6992
Total*	1	1	1	1	1	1	1	1	1	1
Unweighted n	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%
Weighted N	1	1	8	34	44	33	29	43	37	230
	73	306	478	2689	4354	2802	2699	3701	2972	20075

* Totals may not add up due to rounding.

- Denominator of percent is less than 20 unweighted cases.

Table 9. Collapsed MSG and Dunhill Tenure Status by Survey Report of Tenure Status for the Two Sample Groups (Percent of Total).

	Survey Report		
	Own	Rent	Total*
<i>Address Match Between Vendors</i>			
MSG			
Own	71.7%	19.4%	91.1%
Standard error	(2.5)	(2.3)	(1.9)
Unweighted n	427	114	541
Weighted N	37735	10202	47937
Rent	2.5%	6.4%	8.9%
Standard error	(1.0)	(1.5)	(1.9)
Unweighted n	9	36	45
Weighted N	1308	3370	4678
Total*	74.2%	25.8%	100.0%
Standard error	(2.3)	(2.3)	-
Unweighted n	436	150	586
Weighted N	39043	13573	52616
DUNHILL			
Own	57.0%	11.3%	68.3%
Standard error	(2.9)	(1.9)	(2.7)
Unweighted n	351	70	421
Weighted N	30011	5945	35956
Rent	17.2%	14.5%	31.7%
Standard error	(2.7)	(1.9)	(2.7)
Unweighted n	85	80	165
Weighted N	9032	7628	16660
Total*	74.2%	25.8%	100.0%
Standard error	(2.3)	(2.3)	-
Unweighted n	436	150	586

Weighted N	39043	13573	52616
------------	-------	-------	-------

	<u>GDR</u>	<u>SE</u> ⁺	<u>NDR</u>	<u>SE</u> ⁺
<i>Address Match</i>				
MSG	21.9	2.3	16.9	2.7
Dunhill	28.5	3.2	-5.9	3.4
Dunhill-MSG	6.6	3.2	-22.8	2.6
(% of MSG Renters that are correct) – (% of Dunhill Renters that are correct)=				26.3
Standard error				= 7.5
Z				= 3.5

* Totals may not add up due to rounding.

+ SE = Standard error. GDR = Gross Difference Rate; NDR = Net Difference Rate

Table 9. (cont) Collapsed MSG and Dunhill Tenure Status by Survey Report of Tenure Status for the Two Sample Groups (Percent of Total).

	Survey Report		Total*
	Own	Rent	
<i>No Address Match Between Vendors</i>			
MSG			
Own	61.9%	33.9%	95.7%
Standard error	(5.0)	(4.9)	(1.7)
Unweighted n	156	64	220
Weighted N	12421	6797	19217
Rent	3.3%	1.0%	4.3%
Standard error	(1.7)	(0.6)	(1.7)
Unweighted n	7	3	10
Weighted N	662	195	857
Total*	65.2%	34.8%	100.0%
Standard error	(4.9)	(4.9)	-
Unweighted n	163	67	230
Weighted N	13083	6992	20075

	<u>GDR</u>	<u>SE⁺</u>	<u>NDR</u>	<u>SE⁺</u>
<i>No Address Match</i>				
MSG	37.2	5.0	30.6	5.4
<i>(Match) – (No Match)</i>				
MSG	15.3	5.8	13.7	6.4

* Totals may not add up due to rounding.

+ SE = Standard error.

GDR = Gross Difference Rate; NDR = Net Difference Rate

Table 10. MSG Tenure Status by Survey Report of Tenure Status for Mode Groups (Column Percents).

	Rent							Own					Total*
	0	1	2	3	4	5	6	7	8	9	10	11	Total*
<i>In Person</i>													
Own	0%	0.1%	5.8%	6.6%	5.9%	0.2%	8.2%	3.0%	5.7%	1.2%			
Standard error													
Unweighted n	-	-	-	-	(9.8)	(9.3)	(6.0)	(7.4)	(10.5)	(3.6)			
Weighted N	0	1	4	14	21	27	29	33	20	149			
	0	306	1013	3552	4152	4091	4892	5295	3325	26627			
Rent	0%	9.9%	4.3%	3.4%	4.1%	9.8%	1.8%	7.0%	4.3%	8.8%			
Standard error													
Unweighted n	-	-	-	-	(9.8)	(9.3)	(6.0)	(7.4)	(10.5)	(3.6)			
Weighted N	0	3	5	2	10	10	3	15	6	54			
	0	305	1200	551	2147	1737	652	3107	1068	10767			
Total*	0%	00.0%	00.0%	00.0%	00.0%	00.0%	000%	00.0%	00.0%	00.0%			
Unweighted n	0	4	9	16	31	37	32	48	26	203			
Weighted N	0	612	2213	4103	6299	5828	5545	8402	4394	37394			
<i>Telephone</i>													
Own	0.1%	5.4%	7.2%	9.7%	2.8%	4.4%	3.2%	6.2%	6.6%	2.2%			
Standard error													
Unweighted n	-	-	(9.6)	(5.4)	(6.4)	(6.4)	(6.5)	(4.2)	(2.6)	(2.0)			
Weighted N	2	1	8	47	53	50	46	100	143	450			
	85	73											

			493	2461	3143	2623	2892	5165	8564	25499
Rent	7	8	7	4	2	2	2	2	1	2
Standard error	9.9%	4.6%	2.8%	0.3%	7.2%	5.6%	6.8%	3.8%	3.4%	7.8%
Unweighted n	-	-	((((((((
Weighted N	6	5	120	27	23	17	16	28	21	163
	337	403	1319	1659	1176	904	1058	1617	1323	9797
Total*	1	1	1	1	1	1	1	1	1	1
Unweighted n	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%
Weighted N	8	6	28	74	76	67	62	128	164	613
	422	477	1812	4120	4319	3527	3950	6782	9887	35296

* Totals may not add up due to rounding.

- Denominator of percent is less than 20 unweighted cases.

Table 11. Collapsed MSG and Dunhill Tenure Status by Survey Report of Tenure Status for Mode Groups (Percent of Total)*.

	Survey Report		Total**
	Own	Rent	
<i>In-Person</i>			
MSG			
Own	73.0%	17.8%	90.8%
Standard error	(4.2)	(4.1)	(3.6)
Unweighted n	110	24	134
Weighted N	19181	4673	23853
Rent	3.5%	5.7%	9.2%
Standard error	(1.9)	(2.6)	(3.6)
Unweighted n	3	8	11
Weighted N	912	1506	2417
Total**	76.5%	23.5%	100.0%
Standard error	(3.9)	(3.9)	-
Unweighted n	113	32	145
Weighted N	20092	6178	26271
Dunhill			
Own	54.7%	9.8%	64.5%
Standard error	(5.4)	(3.4)	(5.0)
Unweighted n	82	14	96
Weighted N	14364	2567	16931
Rent	21.8%	13.8%	35.6%
Standard error	(5.0)	(3.2)	(5.0)
Unweighted n	31	18	49
Weighted N	5728	3611	9340
Total**	76.5%	23.5%	100.0%
Standard error	(3.9)	(3.9)	-
Unweighted n	113	32	145
Weighted N	20092	6178	26271

	<u>GDR</u>	<u>SE⁺</u>	<u>NDR</u>	<u>SE⁺</u>	
<i>In-Person</i>					
MSG	21.3	4.0	14.3	4.9	
Dunhill	31.6	6.0	-12.0	6.1	
Dunhill-MSG	10.3	5.7	-26.4	4.6	
(% of MSG Renters that are correct) – (% of Dunhill Renters that are correct) =					
23.6					
Standard error					= 13.2
Z					= 1.8

* Excludes units that did not match Dunhill file.

** Totals may not add up due to rounding.;

⁺SE = Standard error, GDR =Gross Difference Rate; NDR = Net Difference Rate

Table 11. (cont) Collapsed MSG and Dunhill Tenure Status by Survey Report of Tenure Status for Mode Groups (Percent of Total)*.

	Survey Report		Total**
	Own	Rent	
<i>Telephone</i>			
MSG			
Own	70.4%	21.0%	91.4%
Standard error	(2.7)	(2.2)	(1.5)
Unweighted n	317	90	407
Weighted N	18554	5530	24084
Rent	1.5%	7.1%	8.6%
Standard error	(0.7)	(1.4)	(1.5)
Unweighted n	6	28	34
Weighted N	397	1864	2261
Total**	71.9%	28.1%	100.0%
Standard error	(2.5)	(2.5)	-
Unweighted n	323	118	441
Weighted N	18951	7394	26345
Dunhill			
Own	59.4%	12.8%	72.2%
Standard error	(2.3)	(1.8)	(2.2)
Unweighted n	269	56	325
Weighted N	15647	3378	19025
Rent	12.5%	15.3%	27.8%
Standard error	(1.8)	(2.0)	(2.2)
Unweighted n	54	62	116
Weighted N	3304	4016	7320
Total**	71.9%	28.1%	100.0%
Standard error	(2.5)	(2.5)	-
Unweighted n	323	118	441
Weighted N	18951	7394	26345

GDR

SE⁺

NDR

SE⁺

Telephone

MSG	22.5	2.5	19.5	2.2	
Dunhill	25.4	2.2	0.3	2.8	
Dunhill-MSG	2.9	2.7	-19.2	2.5	
(% of MSG Renters that are correct) – (% of Dunhill Renters that are correct)					=
27.6					
Standard error					= 8.6
Z					= 3.2

* Excludes units that did not match Dunhill file.

** Totals may not add up due to rounding.

+ SE = Standard error. GDR = Gross Difference Rate; NDR = Net Difference Rate

Table 12. MSG Tenure Status by Survey Report of Tenure Status for Percent Renters on Block (Column Percents).

	Rent											Own											Total*								
	0	1	2	3	4	5	6	7	8	9	10	11	12	0	1	2	3	4	5	6	7	8	9	10	11	12					
0-10%																															
Own	0.0%	0.8%	5.3%	9.6%	3.8%	7.2%	1.0%	6.1%	6.6%	9.7%	0.0%	0.8%	5.3%	9.6%	3.8%	7.2%	1.0%	6.1%	6.6%	9.7%	0.0%	0.8%	5.3%	9.6%	3.8%	7.2%	1.0%	6.1%	6.6%	9.7%	
Standard error	-	-	-	(((((((-	-	-	((((((((-	-	-	(((((((
Unweighted n	1	1	2	30	36	34	46	75	104	329	0	2	8	4	8	9	4	17	14	66	1	3	10	34	44	43	50	92	118	395	
Weighted N	12	73	365	3326	3350	3276	4572	6362	7893	29230	0	165	1076	385	1190	968	455	2003	1225	7466	12	238	1441	3712	4540	4243	5027	8366	9118	36696	
Rent	0%	9.2%	4.7%	0.4%	6.2%	2.8%	.1%	4.0%	3.4%	0.4%	0%	9.2%	4.7%	0.4%	6.2%	2.8%	.1%	4.0%	3.4%	0.4%	0%	9.2%	4.7%	0.4%	6.2%	2.8%	.1%	4.0%	3.4%	0.4%	
Standard error	-	-	-	(((((((-	-	-	((((((((-	-	-	(((((((
Unweighted n	0	2	8	4	8	9	4	17	14	66	0	2	8	4	8	9	4	17	14	66	0	2	8	4	8	9	4	17	14	66	
Weighted N	0	165	1076	385	1190	968	455	2003	1225	7466	0	165	1076	385	1190	968	455	2003	1225	7466	0	165	1076	385	1190	968	455	2003	1225	7466	
Total*	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
Unweighted n	1	3	10	34	44	43	50	92	118	395	1	3	10	34	44	43	50	92	118	395	1	3	10	34	44	43	50	92	118	395	
Weighted N	12	238	1441	3712	4540	4243	5027	8366	9118	36696	12	238	1441	3712	4540	4243	5027	8366	9118	36696	12	238	1441	3712	4540	4243	5027	8366	9118	36696	
11-20%																															
Own	8.3%	0%	4.8%	9.1%	6.4%	7.8%	4.1%	9.0%	2.3%	2.4%	8.3%	0%	4.8%	9.1%	6.4%	7.8%	4.1%	9.0%	2.3%	2.4%	8.3%	0%	4.8%	9.1%	6.4%	7.8%	4.1%	9.0%	2.3%	2.4%	
Standard error	-	-	-	(((((((-	-	-	((((((((-	-	-	(((((((
Unweighted n	1	0	5	18	23	29	18	38	36	168	1	0	5	18	23	29	18	38	36	168	1	0	5	18	23	29	18	38	36	168	
Weighted N	73	0	480	1981	2462	2255	2182	2569	2480	14482	73	0	480	1981	2462	2255	2182	2569	2480	14482	73	0	480	1981	2462	2255	2182	2569	2480	14482	
Rent	1.7%	0.0%	5.2%	0.9%	3.6%	2.2%	5.9%	1.0%	7.8%	7.6%	1.7%	0.0%	5.2%	0.9%	3.6%	2.2%	5.9%	1.0%	7.8%	7.6%	1.7%	0.0%	5.2%	0.9%	3.6%	2.2%	5.9%	1.0%	7.8%	7.6%	
Standard error	-	-	-	(((((((-	-	-	((((((((-	-	-	(((((((
	3			10.0)	8.0)	15.2)	7.5)	8.9)	7.1)	3.5)	3			10.0)	8.0)	15.2)	7.5)	8.9)	7.1)	3.5)	3			10.0)	8.0)	15.2)	7.5)	8.9)	7.1)	3.5)	

Unweighted n	186	1	5	5	12	9	6	15	9	65
Weighted N		73	396	523	759	1070	413	1152	953	5525
Total*	00.0% ¹	00.0% ¹	00.0% ¹	00.0% ¹	00.0% ¹	00.0% ¹	00.0% ¹	00.0% ¹	00.0% ¹	00.0% ¹
Unweighted n	4	1	10	23	35	38	24	53	45	233
Weighted N	259	73	876	2504	3220	3325	2595	3721	3433	20007

* Totals may not add up due to rounding.

- Denominator of percent is less than 20 unweighted cases.

Table 12. (cont.) MSG Tenure Status by Survey Report of Tenure Status for Percent Renters on Block (Column Percents).

	R ent										O wn	T otal*
	0	1	2	3	4	5	6	7	8	9	10	
21-30%												
Own	0	0	5	3	7	8	4	6	8	6	6	
Standard error	-	-	-	((((((((
Unweighted n	0	0	5	9	11	11	5	14	19	74		
Weighted N	0	0	660	490	810	975	559	819	1198	5511		
Rent	0	1	4	6	2	1	5	3	1	3	3	
Standard error	-	-	-	((((((((
Unweighted n	0	5	4	12	6	5	8	5	3	48		
Weighted N	0	470	545	807	295	124	623	471	170	3505		
Total*	1	1	1	1	1	1	1	1	1	1	1	
Unweighted n	0	5	9	21	17	16	13	19	22	122		
Weighted N	0	470	1205	1298	1105	1099	1182	1290	1368	9017		
31-40%												
Own	0	1	0	3	3	3	6	3	8	4	4	
Standard error	-	-	-	-	-	-	-	-	-	((
Unweighted n	0	1	0	4	4	3	6	6	4	28		
Weighted N	0	306	0	216	674	208	471	710	318	2903		
Rent	1	0	1	6	6	6	3	6	1	5	5	
Standard error	-	-	-	-	-	-	-	-	-	((
	3	0	8	8	7	4	1	6	1	7.1)	7.1)	

Unweighted n	152	0	502	493	1079	480	219	1098	44	38
Weighted N										4067
Total*	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%
Unweighted n	3	1	8	12	11	7	7	12	5	66
Weighted N	152	306	502	709	1753	688	690	1807	363	6971

* Totals may not add up due to rounding.

- Denominator of percent is less than 20 unweighted cases.

Table 13. Collapsed MSG and Dunhill Tenure Status by Survey Report for Percent Renters on Block (Percent of Total)*.

	Survey Report		Total**
	Own	Rent	
0% - 10% Renter			
MSG			
Own	81.3%	13.2%	94.5%
Standard error	(2.7)	(2.4)	(1.9)
Unweighted n	250	36	286
Weighted N	22115	3591	25707
Rent	1.4%	4.1%	5.5%
Standard error	(1.1)	(1.8)	(1.9)
Unweighted n	3	8	11
Weighted N	377	1119	1496
Total**	82.7%	17.3%	100.0%
Standard error	(2.9)	(2.9)	-
Unweighted n	253	44	297
Weighted N	22492	4711	27203
Dunhill			
Own	65.3%	8.4%	73.7%
Standard error	(3.9)	(2.3)	(3.6)
Unweighted n	204	25	229
Weighted N	17753	2290	20043
Rent	17.4%	8.9%	26.3%
Standard error	(3.7)	(2.5)	(3.6)
Unweighted n	49	19	68
Weighted N	4739	2421	7160
Total**	82.7%	17.3%	100.0%
Standard error	(2.9)	(2.9)	-
Unweighted n	253	44	297
Weighted N	22492	4711	27203

	<u>GDR</u>	<u>SE</u> ⁺	<u>NDR</u>	<u>SE</u> ⁺
<i>0% - 10% Renter</i>				
MSG	14.6	2.3	11.8	2.9
Dunhill	25.8	4.4	-9.0	4.4
Dunhill- MSG	11.2	4.5	-20.8	3.6
(% of MSG Renters that are correct) – (% of Dunhill Renters that are correct)	=			
41.0				
Standard error				= 17.9
Z				= 2.3

* Excludes units that did not match Dunhill file.

** Totals may not add up due to rounding.

⁺ SE = Standard error. GDR = Gross Difference Rate; NDR = Net Difference Rate.

Table 13. (cont.) Collapsed MSG and Dunhill Tenure Status by Survey Report for Percent Renters on Block (Percent of Total)*.

	Survey Report		Total**
	Own	Rent	
11% - 20% Renter			
MSG			
Own	71.5%	21.5%	93.0%
Standard error	(4.6)	(4.0)	(2.5)
Unweighted n	114	36	150
Weighted N	10761	3236	13997
Rent	2.7%	4.4%	7.0%
Standard error	(1.8)	(1.5)	(2.5)
Unweighted n	3	9	12
Weighted N	402	655	1057
Total**	74.2%	25.9%	100.0%
Standard error	(4.2)	(4.2)	-
Unweighted n	117	45	162
Weighted N	11163	3891	15054
Dunhill			
Own	53.6%	11.2%	64.8%
Standard error	(5.0)	(2.9)	(6.1)
Unweighted n	93	21	114
Weighted N	8069	1685	9755
Rent	20.6%	14.7%	35.2%
Standard error	(5.6)	(3.0)	(6.1)
Unweighted n	24	24	48
Weighted N	3093	2206	5299
Total**	74.2%	25.9%	100.0%
Standard error	(4.2)	(4.2)	-
Unweighted n	117	45	162
Weighted N	11163	3891	15054

	<u>GDR</u>	<u>SE</u> ⁺	<u>NDR</u>	<u>SE</u> ⁺
<i>11% - 20% Renter</i>				
MSG	24.2	4.4	18.8	4.4
Dunhill	31.7	4.3	-9.4	7.6
Dunhill- MSG	7.6	6.5	-28.2	6.5
(% of MSG Renters that are correct) – (% of Dunhill Renters that are correct)				=
20.4				
Standard error				= 21.3
Z				= 1.0

* Excludes units that did not match Dunhill file.

** Totals may not add up due to rounding.

⁺ SE = Standard error. GDR = Gross Difference Rate; NDR = Net Difference Rate.

Table 13. (cont.) Collapsed MSG and Dunhill Tenure Status by Survey Report for Percent Renters on Block (Percent of Total)*.

	Survey Report		Total**
	Own	Rent	
21% - 30% Renter			
MSG			
Own	49.4%	25.6%	75.0%
Standard error	(9.6)	(6.4)	(12.3)
Unweighted n	44	25	69
Weighted N	3048	1581	4629
Rent	8.6%	16.4%	25.0%
Standard error	(6.1)	(8.0)	(12.3)
Unweighted n	3	9	12
Weighted N	530	1015	1545
Total**	58.0%	42.1%	100.0%
Standard error	(6.8)	(6.8)	-
Unweighted n	47	34	81
Weighted N	3578	2596	6174
Dunhill			
Own	41.7%	15.4%	57.1%
Standard error	(9.3)	(4.7)	(10.0)
Unweighted n	38	15	53
Weighted N	2575	952	3527
Rent	16.2%	26.6%	42.9%
Standard error	(5.7)	(6.5)	(10.0)
Unweighted n	9	19	28
Weighted N	1003	1645	2647
Total**	58.0%	42.1%	100.0%
Standard error	(6.8)	(6.8)	-
Unweighted n	47	34	81
Weighted N	3578	2596	6174

	<u>GDR</u>	<u>SE</u> ⁺	<u>NDR</u>	<u>SE</u> ⁺
<i>21% - 30% Renter</i>				
MSG	34.2	6.1	17.010.9	
Dunhill	31.7	6.0	-0.88.6	
Dunhill-MSG	-2.5	5.3	-17.96.0	
(% of MSG Renters that are correct) – (% of Dunhill Renters that are correct)				=
3.6				
Standard error				= 14.2
Z				= 0.3

* Excludes units that did not match Dunhill file.

** Totals may not add up due to rounding.

⁺ SE = Standard error. GDR = Gross Difference Rate; NDR = Net Difference Rate.

Table 13. (cont.) Collapsed MSG and Dunhill Tenure Status by Survey Report for Percent Renters on Block (Percent of Total)*.

	Survey Report		Total**
	Own	Rent	
31% - 40% Renter			
MSG			
Own	43.3%	42.9%	86.1%
Standard error	(10.1)	(10.4)	(4.8)
Unweighted n	19	17	36
Weighted N	1811	1794	3605
Rent	0%	13.9%	13.9%
Standard error	-	(4.8)	(4.8)
Unweighted n	0	10	10
Weighted N	0	581	581
Total**	43.3%	56.7%	100.0%
Standard error	(10.1)	(10.1)	-
Unweighted n	19	27	46
Weighted N	1811	2374	4185
Dunhill			
Own	38.6%	24.3%	62.9%
Standard error	(10.1)	(11.7)	(8.1)
Unweighted n	16	9	25
Weighted N	1614	1018	2631
Rent	4.7%	32.4%	37.1%
Standard error	(2.8)	(7.6)	(8.1)
Unweighted n	3	18	21
Weighted N	197	1357	1554
Total**	43.3%	56.7%	100.0%
Standard error	(10.1)	(10.1)	-
Unweighted n	19	27	46
Weighted N	1811	2374	4185

	<u>GDR</u>	<u>SE</u> ⁺	<u>NDR</u>	<u>SE</u> ⁺
<i>31% - 40% Renter</i>				
MSG	42.9	10.4	42.9	10.4
Dunhill	29.0	11.7	19.6	12.3
Dunhill-MSG	-13.8	7.1	-23.3	7.4
(% of MSG Renters that are correct) – (% of Dunhill Renters that are correct)=				12.7
Standard error				= 7.5
Z				= 1.7

* Excludes units that did not match Dunhill file. ** Totals may not add up due to rounding.

⁺ SE = Standard error. GDR = Gross Difference Rate; NDR = Net Difference Rate.

Table 14. MSG Tenure Status by Survey Report of Tenure Status for Each County (Column Percent).

	R ent											O wn											Total*								
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Total*									
Baltimore County, MD																															
Own	5.0%	3.3%	7.0%	5.9%	9.7%	2.7%	2.4%	1.8%	2.5%	0.8%	5.0%	3.3%	7.0%	5.9%	9.7%	2.7%	2.4%	1.8%	2.5%	0.8%	5.0%	3.3%	7.0%	5.9%	9.7%	2.7%	2.4%	1.8%	2.5%	0.8%	
Standard error	-	-	-	(((((((-	-	-	((((((((-	-	-	(((((((
Unweighted n	1	2	5	22	29	22	20	40	62	203	3	3	7	12	11	6	8	20	12	82	4	5	12	34	40	28	28	60	74	285	
Weighted N	73	380	748	4036	4779	3269	3128	5107	6626	28146	220	220	846	1284	2076	1225	1190	3157	1410	82	293	600	1594	5320	6854	4494	4318	8264	8036	39773	
Rent	5.0%	6.7%	3.1%	4.1%	0.3%	7.3%	7.6%	8.2%	7.5%	9.2%	5.0%	6.7%	3.1%	4.1%	0.3%	7.3%	7.6%	8.2%	7.5%	9.2%	5.0%	6.7%	3.1%	4.1%	0.3%	7.3%	7.6%	8.2%	7.5%	9.2%	
Standard error	-	-	-	(((((((-	-	-	((((((((-	-	-	(((((((
Unweighted n	3	3	7	12	11	6	8	20	12	82	3	3	7	12	11	6	8	20	12	82	3	3	7	12	11	6	8	20	12	82	
Weighted N	220	220	846	1284	2076	1225	1190	3157	1410	82	220	220	846	1284	2076	1225	1190	3157	1410	82	293	600	1594	5320	6854	4494	4318	8264	8036	39773	
Total*	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	
Unweighted n	4	5	12	34	40	28	28	60	74	285	4	5	12	34	40	28	28	60	74	285	4	5	12	34	40	28	28	60	74	285	
Weighted N	293	600	1594	5320	6854	4494	4318	8264	8036	39773	293	600	1594	5320	6854	4494	4318	8264	8036	39773	293	600	1594	5320	6854	4494	4318	8264	8036	39773	
Howard County, MD																															
Own	0	0	3	5	6	6	9	7	8	6	0	0	3	5	6	6	9	7	8	6	0	0	3	5	6	6	9	7	8	6	
Standard error	-	%	0.9%	0.0%	4.6%	3.1%	2.2%	7.3%	4.6%	7.2%	-	%	0.9%	0.0%	4.6%	3.1%	2.2%	7.3%	4.6%	7.2%	-	%	0.9%	0.0%	4.6%	3.1%	2.2%	7.3%	4.6%	7.2%	
Unweighted n	0	-	-	-	-	-	-	-	((0	-	-	-	-	-	-	-	-	((0	-	-	-	-	-	-	-	((
Weighted N	0	0	3	6	3	7	14	15	22	70	0	0	3	6	3	7	14	15	22	70	0	0	3	6	3	7	14	15	22	70	
			566	550	622	1403	2194	1402	1524	8261			566	550	622	1403	2194	1402	1524	8261			566	550	622	1403	2194	1402	1524	8261	
Rent	0	1	6	5	3	3	7	2	1	3	0	1	6	5	3	3	7	2	1	3	0	1	6	5	3	3	7	2	1	3	

Standard error	-	00.0%	9.1%	0.0%	5.4%	6.9%	.9%	2.7%	5.4%	2.8%
Unweighted n	0	-	-	-	-	-	-	-	((
Weighted N	0	2	9	6	2	5	2	3	6.0)	2.5)
		183	1268	550	340	822	187	412	4	33
									277	4040
Total*	1	1	1	1	1	1	1	1	1	1
Unweighted n	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%
Weighted N	0	2	12	12	5	12	16	18	26	103
	0	183	1834	1100	962	2226	2381	1814	1801	12300

* Totals may not add up due to rounding.

- Denominator of percent is less than 20 unweighted cases.

Table 14. (cont) MSG Tenure Status by Survey Report of Tenure Status for Each County (Column Percent).

	Rent							Own				Total*
	0	1	2	3	4	5	6	7	8	9	Total*	
Queen Anne's County, MD												
Own	1	0	1	8	5	6	5	8	8	7	7	
Standard error	00.0%	-	00.0%	0.0%	0.0%	8.8%	8.3%	2.1%	8.9%	4.2%	(
Unweighted n	-	0	-	-	-	-	-	(-	((
Weighted N	1	0	1	4	6	11	7	6.4)	16	69	5.2)	
	12	12	47	91	167	200	23	486	338	1353		
Rent	0	0	0	2	5	3	4	1	1	2	2	
Standard error	%	-	%	0.0%	0.0%	1.3%	1.7%	7.9%	1.1%	5.8%	(
Unweighted n	-	0	-	-	-	-	-	(-	((
Weighted N	0	0	0	1	6	5	5	6.4)	2	24	5.2)	
	0	0	0	12	91	76	143	5	42	470		
								106				
Total*	1	1	1	1	1	1	1	1	1	1	1	
Unweighted n	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	
Weighted N	1	0	1	5	12	16	12	28	18	93		
	12	0	12	58	182	243	342	592	380	1822		
Hanover County, VA												
Own	0	0	0	6	7	7	1	7	7	7	7	
Standard error	-	%	%	0.0%	5.7%	4.3%	00.0%	9.8%	6.0%	0.7%	(
Unweighted n	0	-	-	-	-	-	-	(-	((
Weighted N	0	0	0	4	6	8	6	8.3)	14	56	6.9)	
				135	251	295	513	18	793	3406		
								1419				
Rent	0	1	1	4	2	2	0	2	2	2	2	

Standard error	-	00.0%	00.0%	0.0%	4.3%	5.7%	%	0.3%	4.0%	9.3%
Unweighted n	0	-	-	-	-	-	-	(-	(
Weighted N	0	3	5	3	1	3	0	8.3)	3	6.9)
		305	222	90	80	102	0	5	251	23
								360		1411
Total*	1	1	1	1	1	1	1	1	1	1
Unweighted n	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%
Weighted N	0	3	5	7	7	11	6	23	17	79
	0	305	222	226	331	397	513	1780	1044	4818

* Totals may not add up due to rounding.

- Denominator of percent is less than 20 unweighted cases.

Table 14. (cont) MSG Tenure Status by Survey Report of Tenure Status for Each County (Column Percent).

	Rent							Own				Total*
	0	1	2	3	4	5	6	7	8	9	10	Total*
<i>Henrico County, VA</i>												
Own	0	0	4	8	6	7	9	7	8	7	7	7
Standard error	%	-	9.4%	2.0%	7.9%	9.2%	0.2%	4.8%	6.4%	8.4%		
Unweighted n	-	0	-	(((((((((
Weighted N	0	0	3	25	30	29	28	37	49	201		1
	0	0	179	1245	1553	1579	1749	2046	2608	0960		1
Rent	1	0	5	1	3	2	9	2	1	2	2	2
Standard error	00.0%	-	0.6%	8.0%	2.1%	0.8%	.9%	5.2%	3.6%	1.6%		
Unweighted n	-	0	-	(((((((((
Weighted N	3	0	4	7	13	8	4	10	6	55		
	117	0	184	274	735	416	191	689	412	3017		
Total*	1	1	1	1	1	1	1	1	1	1	1	1
Unweighted n	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%	00.0%
Weighted N	3	0	7	32	43	37	32	47	55	256		1
	117	0	363	1519	2288	1995	1940	734	3020	3977		1

* Totals may not add up due to rounding.

- Denominator of percent is less than 20 unweighted cases.

Table 15. Collapsed MSG and Dunhill Tenure Status by Survey Report of Tenure Status for Each County (Percent of Total)*.

	Survey Report		Total**
	Own	Rent	
<i>Baltimore County, MD</i>			
MSG			
Own	71.5%	21.6%	93.1%
Standard error	(3.8)	(3.7)	(2.0)
Unweighted n	156	47	203
Weighted N	20970	6332	27302
Rent	2.8%	4.1%	6.9%
Standard error	(1.3)	(1.7)	(2.0)
Unweighted n	6	12	18
Weighted N	822	1212	2034
Total**	74.3%	25.7%	100.0%
Standard error	(3.7)	(3.7)	-
Unweighted n	162	59	221
Weighted N	21792	7544	29336
Dunhill			
Own	57.7%	11.3%	69.0%
Standard error	(4.7)	(3.0)	(4.0)
Unweighted n	129	28	157
Weighted N	16913	3325	20237
Rent	16.6%	14.4%	31.0%
Standard error	(4.4)	(2.9)	(4.0)
Unweighted n	33	31	64
Weighted N	4879	4219	9098
Total**	74.3%	25.7%	100.0%
Standard error	(3.7)	(3.7)	-
Unweighted n	162	59	221

Weighted N	21792	7544	29336
------------	-------	------	-------

		<u>GDR</u>	<u>SE</u> ⁺	<u>NDR</u>	<u>SE</u> ⁺
<i>Baltimore County, MD</i>					
4.0	MSG	24.4	3.8	18.8	
5.2	Dunhill	28.0	5.4	-5.3	

* Excludes units that did not match Dunhill file.

** Totals may not add up due to rounding.

⁺ SE = Standard error.

GDR = Gross Difference Rate; NDR = Net Difference Rate

Table 15. (cont) Collapsed MSG and Dunhill Tenure Status by Survey Report of Tenure Status for Each County (Percent of Total)*.

	Survey Report		
	Own	Rent	Total**
Howard County, MD			
MSG			
Own	66.1%	14.7%	80.8%
Standard error	(6.1)	(4.3)	(6.8)
Unweighted n	54	14	68
Weighted N	6305	1404	7710
Rent	5.0%	14.3%	19.2%
Standard error	(4.1)	(3.7)	(6.8)
Unweighted n	2	10	12
Weighted N	475	1360	1834
Total**	71.0%	29.0%	100.0%
Standard error	(5.0)	(5.0)	-
Unweighted n	56	24	80
Weighted N	6780	2764	9544
Dunhill			
Own	48.7%	10.9%	59.6%
Standard error	(5.8)	(5.3)	(8.6)
Unweighted n	43	9	52
Weighted N	4647	1039	5686
Rent	22.4%	18.1%	40.4%
Standard error	(6.0)	(5.2)	(8.6)
Unweighted n	13	15	28
Weighted N	2133	1725	3858
Total**	71.0%	29.0%	100.0%
Standard error	(5.0)	(5.0)	-
Unweighted n	56	24	80
Weighted N			

	6780	2764	9544
--	------	------	------

		<u>GDR</u>	<u>SE</u> ⁺	<u>NDR</u>	<u>SE</u> ⁺
<i>Howard County, MD</i>					
7.5	MSG	19.7	3.8	9.7	
10.2	Dunhill	33.2	5.0	-11.5	

* Excludes units that did not match Dunhill file.

** Totals may not add up due to rounding.

+ SE = Standard error.

GDR = Gross Difference Rate; NDR = Net Difference Rate

Table 15. (cont) Collapsed MSG and Dunhill Tenure Status by Survey Report of Tenure Status for Each County (Percent of Total)*.

	Survey Report		
	Own	Rent	Total**
Queen Anne's County, MD			
MSG			
Own	78.5%	20.6%	99.1%
Standard error	(6.4)	(6.5)	(0.9)
Unweighted n	52	14	66
Weighted N	1056	278	1333
Rent	0.9%	0%	0.9%
Standard error	(0.9)		(0.9)
Unweighted n	1	0	1
Weighted N	12	0	12
Total**	79.4%	20.6%	100.0%
Standard error	(6.5)	(6.5)	-
Unweighted n	53	14	67
Weighted N	1067	278	1345
Dunhill			
Own	67.4%	11.3%	78.6%
Standard error	(6.7)	(4.0)	(6.1)
Unweighted n	45	8	53
Weighted N	906	151	1058
Rent	12.0%	9.4%	21.4%
Standard error	(3.8)	(4.5)	(6.1)
Unweighted n	8	6	14
Weighted N	161	126	287
Total**	79.4%	20.6%	100.0%
Standard error	(6.5)	(6.5)	-
Unweighted n	53	14	67
Weighted N	1067	278	1345

	<u>GDR</u>	<u>SE</u> ⁺	<u>NDR</u>	<u>SE</u> ⁺
<i>Queen Anne's County, MD</i>				
MSG	21.5	6.4	19.86.6	
Dunhill	23.2	4.1	-0.76.6	

* Excludes units that did not match Dunhill file.

** Totals may not add up due to rounding.

+ SE = Standard error.

GDR = Gross Difference Rate; NDR = Net Difference Rate

Table 15. (cont) Collapsed MSG and Dunhill Tenure Status by Survey Report of Tenure Status for Each County (Percent of Total)*.

	Survey Report		
	Own	Rent	Total**
<i>Hanover County, VA</i>			
MSG			
Own	74.5%	8.1%	82.5%
Standard error	(12.6)	(4.0)	(13.9)
Unweighted n	36	5	41
Weighted N	2116	229	2345
Rent	0%	17.5%	17.5%
Standard error	-	(13.9)	(13.9)
Unweighted n	0	7	7
Weighted N	0	497	497
Total**	74.5%	25.6%	100.0%
Standard error	(12.6)	(12.6)	-
Unweighted n	36	12	48
Weighted N	2116	726	2842
Dunhill			
Own	67.1%	10.4%	77.5%
Standard error	(13.7)	(3.9)	(11.6)
Unweighted n	32	5	37
Weighted N	1906	297	2203
Rent	7.4%	15.1%	22.5%
Standard error	(2.6)	(10.2)	(11.6)
Unweighted n	4	7	11
Weighted N	209	429	639
Total**	74.5%	25.6%	100.0%
Standard error	(12.6)	(12.6)	-
Unweighted n	36	12	48
Weighted N	2116	726	2842

GDR SE⁺ NDR SE⁺

Hanover County, VA

MSG	8.1	4.0	8.14.0
Dunhill	17.8	4.6	3.14.9

* Excludes units that did not match Dunhill file.

** Totals may not add up due to rounding.

+ SE = Standard error.

GDR = Gross Difference Rate; NDR = Net Difference Rate

Table 15. (cont) Collapsed MSG and Dunhill Tenure Status by Survey Report of Tenure Status for Each County (Percent of Total)*.

	Survey Report		
	Own	Rent	Total**
Henrico County, VA			
MSG			
Own	76.3%	20.5%	96.9%
Standard error	(3.7)	(3.9)	(1.1)
Unweighted n	129	34	163
Weighted N	7288	1960	9248
Rent	0%	3.2%	3.2%
Standard error		(1.1)	(1.1)
Unweighted n	0	7	7
Weighted N	0	301	301
Total**	76.3%	23.7%	100.0%
Standard error	(3.7)	(3.7)	-
Unweighted n	129	41	170
Weighted N	7288	2261	9549
Dunhill			
Own	59.1%	11.9%	70.9%
Standard error	(3.6)	(2.6)	(2.9)
Unweighted n	102	20	122
Weighted N	5639	1133	6772
Rent	17.3%	11.8%	29.1%
Standard error	(3.2)	(2.5)	(2.9)
Unweighted n	27	21	48
Weighted N	1650	1128	2778
Total**	76.3%	23.7%	100.0%
Standard error	(3.7)	(3.7)	-
Unweighted n	129	41	170
Weighted N	7288	2261	9549

GDR SE⁺ NDR SE

Henrico County, VA

MSG	20.5	3.9	20.5	3.9
Dunhill	29.1	3.7	-5.4	4.5

* Excludes units that did not match Dunhill file.

** Totals may not add up due to rounding.

+ SE = Standard error.

GDR = Gross Difference Rate; NDR = Net Difference Rate

APPENDIX A
SUPPLEMENTARY TABLES

Table A1. Collapsed MSG Tenure Status by Survey Report of Tenure Status for Mode Groups (Percent of Total).

	Survey Report		Total*
	Own	Rent	
<i>In-Person</i>			
MSG			
Own	67.7%	24.8%	92.5%
Standard error	(3.8)	(3.7)	(2.6)
Unweighted n	144	46	190
Weighted N	25308	9262	34570
Rent	3.5%	4.0%	7.6%
Standard error	(1.5)	(1.9)	(2.6)
Unweighted n	5	8	13
Weighted N	1319	1506	2825
Total*	71.2%	28.8%	100.0%
Standard error	(3.6)	(3.6)	-
Unweighted n	149	54	203
Weighted N	26627	10767	37394
<i>Telephone</i>			
MSG			
Own	70.4%	21.9%	92.3%
Standard error	(2.1)	(1.9)	(1.1)
Unweighted n	439	132	571
Weighted N	24848	7737	32585
Rent	1.9%	5.8%	7.7%
Standard error	(0.6)	(1.0)	(1.1)
Unweighted n	11	31	42
Weighted N	651	2059	2711
Total*	72.2%	27.8%	100.0%
Standard error	(2.0)	(2.0)	-
Unweighted n	450	163	613
Weighted N	25499	9797	35296

		<u>GDR</u>	<u>SE</u> ⁺	<u>NDR</u>	<u>SE</u> ⁺
<i>In-Person</i>					
4.2	MSG	28.3	3.8	21.2	
	<i>Telephone</i>				
2.0	MSG	23.8	2.1	20.1	

* Totals may not add up due to rounding.

+ SE = standard error.

GDR = Gross Difference Rate; NDR = Net Difference Rate

Table A2. Collapsed MSG Tenure Status by Survey Report of Tenure Status for Percent Renters on Block (Percent of Total).

	Survey Report		
	Own	Rent	Total*
<i>0% - 10% Renter</i>			
MSG			
Own	78.4%	17.0%	95.4%
Standard error	(3.6)	(3.5)	(1.5)
Unweighted n	325	56	381
Weighted N	28780	6225	35005
Rent	1.2%	3.4%	4.6%
Standard error	(0.8)	(1.3)	(1.5)
Unweighted n	4	10	14
Weighted N	450	1241	1691
Total*	79.7%	20.4%	100.0%
Standard error	(3.8)	(3.8)	-
Unweighted n	329	66	395
Weighted N	29230	7466	36696
<i>11% - 20% Renter</i>			
MSG			
Own	69.6%	24.3%	94.0%
Standard error	(3.9)	(3.4)	(1.9)
Unweighted n	162	56	218
Weighted N	13928	4870	18798
Rent	2.8%	3.3%	6.0%
Standard error	(1.5)	(1.1)	(1.9)
Unweighted n	6	9	15
Weighted N	553	655	1209
Total*	72.4%	27.6%	100.0%
	(3.5)	(3.5)	-

Standard error	168	65	233
Unweighted n	14482	5525	20007
Weighted N			

		<u>GDR</u>	<u>SE⁺</u>	<u>NDR</u>	<u>SE⁺</u>
<i>0% - 10% Renter</i>					
3.8	MSG	18.2	3.4	15.7	
<i>11% - 20% Renter</i>					
3.6	MSG	27.1	3.7	21.6	

* Totals may not add up due to rounding.

+ SE = standard error (based on unweighted sample size.)

GDR = Gross Difference Rate; NDR = Net Difference Rate

Table A2. (cont) Collapsed MSG Tenure Status by Survey Report of Tenure Status for Percent Renters on Block (Percent of Total).

	Survey Report		
	Own	Rent	Total*
21% - 30% Renter			
MSG			
Own	53.8%	27.6%	81.4%
Standard error	(6.7)	(5.3)	(8.8)
Unweighted n	69	39	108
Weighted N	4851	2490	7341
Rent	7.3%	11.3%	18.6%
Standard error	(4.3)	(5.8)	(8.8)
Unweighted n	5	9	14
Weighted N	660	1015	1675
Total*	61.1%	38.9%	100.0%
Standard error	(4.9)	(4.9)	-
Unweighted n	74	48	122
Weighted N	5511	3505	9017
31% - 40% Renter			
MSG			
Own	37.3%	49.0%	86.2%
Standard error	(6.5)	(7.7)	(5.4)
Unweighted n	27	27	54
Weighted N	2597	3414	6011
Rent	4.4%	9.4%	13.8%
Standard error	(4.4)	(2.9)	(5.4)
Unweighted n	1	11	12
Weighted N	306	654	960
Total*	41.7%	58.4%	100.0%
Standard error	(7.1)	(7.1)	-
Unweighted n	28	38	66
Weighted N	2903	4067	6971

		<u>GDR</u>	<u>SE</u> ⁺	<u>NDR</u>	<u>SE</u> ⁺
	<i>21% - 30% Renter</i>				
8.3	MSG	34.9	4.9	20.3	
	<i>31% - 40% Renter</i>				
7.7	MSG	49.0	7.7	49.0	

* Totals may not add up due to rounding.

+ SE = standard error (based on unweighted sample size.)

GDR = Gross Difference Rate; NDR = Net Difference Rate

Table A3. Collapsed MSG Tenure Status by Survey Report of Tenure Status for Each County (Percent of Total).

	Survey Report		Total*
	Own	Rent	
<i>Baltimore County, MD</i>			
MSG			
Own	67.8%	26.0%	93.8%
Standard error	(3.7)	(3.4)	(1.6)
Unweighted n	195	69	264
Weighted N	26945	10341	37286
Rent	3.0%	3.2%	6.3%
Standard error	(1.2)	(1.2)	(1.6)
Unweighted n	8	13	21
Weighted N	1201	1285	2487
Total*	70.8%	29.2%	100.0%
Standard error	(3.5)	(3.5)	-
Unweighted n	203	82	285
Weighted N	28146	11627	39773
<i>Howard County, MD</i>			
MSG			
Own	62.6%	21.0%	83.6%
Standard error	(2.9)	(4.4)	(6.1)
Unweighted n	67	22	89
Weighted N	7694	2588	10283
Rent	4.6%	11.8%	16.4%
Standard error	(3.3)	(3.6)	(6.1)
Unweighted n	3	11	14
Weighted N	566	1451	2018
Total*	67.2%	32.8%	100.0%
Standard error	(2.5)	(2.5)	-
Unweighted n	70	33	103
Weighted N	8261	4040	12300

		<u>GDR</u>	<u>SE</u> ⁺	<u>NDR</u>	<u>SE</u> ⁺
<i>Baltimore County, MD</i>					
3.5	MSG	29.0	3.7	23.0	
<i>Howard County, MD</i>					
7.2	MSG	25.7	2.9	16.4	

* Totals may not add up due to rounding.

+ SE = standard error based on unweighted sample size.

GDR = Gross Difference Rate; NDR = Net Difference Rate

Table A3. (cont) Collapsed MSG Tenure Status by Survey Report of Tenure Status for Each County (Percent of Total).

	Survey Report		
	Own	Rent	Total*
<i>Queen Anne's County, MD</i>			
MSG			
Own	73.0%	25.8%	98.7%
Standard error	(5.2)	(5.2)	(0.9)
Unweighted n	67	24	91
Weighted N	1329	470	1799
Rent	1.3%	0%	1.3%
Standard error	(0.9)	-	(0.9)
Unweighted n	2	0	2
Weighted N	23	0	23
Total*	74.2%	25.8%	100.0%
Standard error	(5.2)	(5.2)	-
Unweighted n	69	24	93
Weighted N	1353	470	1822
<i>Hanover County, VA</i>			
MSG			
Own	70.7%	18.4%	89.1%
Standard error	(6.9)	(7.6)	(8.7)
Unweighted n	56	15	71
Weighted N	3406	884	4290
Rent	0%	10.9%	10.9%
Standard error	-	(8.7)	(8.7)
Unweighted n	0	8	8
Weighted N	0	527	527
Total*	70.7%	29.3%	100.0%
Standard error	(6.9)	(6.9)	-
Unweighted n	56	23	79
Weighted N	3406	1411	4818

		<u>GDR</u>	<u>SE</u> ⁺	<u>NDR</u>	<u>SE</u> ⁺
<i>Queen Anne's County, MD</i>					
	MSG	27.1	5.2	24.5	
5.2					
<i>Hanover County, VA</i>					
	MSG	18.4	7.6	18.4	
7.6					

* Totals may not add up due to rounding.

+ SE = standard error based on unweighted sample size.

GDR = Gross Difference Rate; NDR = Net Difference Rate

Table A3. (cont) Collapsed MSG Tenure Status by Survey Report of Tenure Status for Each County (Percent of Total).

	Survey Report		
	Own	Rent	Total*
<i>Henrico County, VA</i>			
MSG			
Own	77.1%	19.4%	96.6%
Standard error	(3.1)	(3.0)	(1.1)
Unweighted n	198	48	246
Weighted N	10781	2716	13497
Rent	1.3%	2.2%	3.4%
Standard error	(0.8)	(0.8)	(1.1)
Unweighted n	3	7	10
Weighted N	179	301	480
Total*	78.4%	21.6%	100.0%
Standard error	(2.8)	(2.8)	-
Unweighted n	201	55	256
Weighted N	10960	3017	13977

GDR SE⁺ NDR SE⁺

Henrico County, VA

MSG 20.7 3.2 18.2 3.0

* Totals may not add up due to rounding.

+ SE = Standard error based on unweighted sample size.

GDR = Gross Difference Rate; NDR = Net Difference Rate

APPENDIX B
TELEPHONE QUESTIONNAIRE

**CPI HOUSING TENURE SURVEY
TELEPHONE VERSION**

Hello, my name is [NAME] and I am calling for the Bureau of Labor Statistics, an agency of the U.S. government. We would like to conduct a 3 minute survey with an adult in this household.

[IF NEEDED: We recently sent you a letter about this study, which is examining ways to improve how the government measures housing costs.]

Before I begin, I need to verify your address.

1. Is this [RECITE ADDRESS FROM ASSIGNMENT LABEL]?

- YES
- NO GO TO END

2. Is this address for a business or a residence?

- RESIDENCE
- BUSINESS GO TO END

3. May I please speak with someone who is at least 18 years old and who lives at this address?

- YES
- NO GO TO END

[OPTIONAL: I'd like to speak with someone who is at least 18 years old and who lives at this address. Would that be you?]

The purpose of this study is to help improve the way the government collects information about housing costs.

I work for a research company called Westat. We are conducting this study for the Bureau of Labor Statistics. Westat and the Bureau of Labor Statistics will use the information you provide for statistical purposes only and will hold the information in confidence to the full extent permitted by law.

I have just a few questions to ask you about your home.

4. First, is this house or apartment owned or being bought by you or someone in your household?

- YES GO TO Q7
- NO

5. Is this house or apartment being rented by you or someone in your household?

- YES
- NO GO TO END

6. How much is your current monthly rent?

- \$ _____ GO TO END
- DON'T KNOW GO TO END
- REFUSED GO TO END

7. If your home were to be rented out, about how much would it rent for per month?

- \$ _____
- DON'T KNOW
- REFUSED

8. What is the least you would accept in rent?

- \$ _____
- DON'T KNOW
- REFUSED

9. Do you currently make a mortgage payment?

- YES
- NO GO TO END

10. What is your mortgage payment each month?

- \$ _____
- DON'T KNOW
- REFUSED

END: Those are all the questions I have for you today. Thank you very much for your participation.

APPENDIX C
IN-PERSON QUESTIONNAIRE

**CPI HOUSING TENURE SURVEY
IN-PERSON VERSION**

Hello, my name is [NAME] and I'm here for the Bureau of Labor Statistics, an agency of the U.S. government. We would like to conduct a 3 minute survey with an adult in this household.

[IF NEEDED: We recently sent you a letter about this study, which is examining ways to improve how the government measures housing costs.]

1. May I please speak with someone who is at least 18 years old and who lives here?

- YES
 NO GO TO END

[OPTIONAL: I'd like to speak with someone who is at least 18 years old and who lives here. Would that be you?]

The purpose of this study is to help improve the way the government collects information about housing costs.

I work for a research company called Westat. We are conducting this study for the Bureau of Labor Statistics. Westat and the Bureau of Labor Statistics will use the information you provide for statistical purposes only and will hold the information in confidence to the full extent permitted by law.

I have just a few questions to ask you about your home.

2. First, is this house or apartment owned or being bought by you or someone in your household?

- YES GO TO Q5

NO

3. Is this house or apartment being rented by you or someone in your household?

YES

NO GO TO END

4. How much is your current monthly rent?

\$ _____
 DON'T KNOW
 REFUSED

GO TO END
GO TO END
GO TO END

5. If your home were to be rented out, about how much would it rent for per month?

- \$ _____
 DON'T KNOW
 REFUSED

6. What is the least you would accept in rent?

- \$ _____
 DON'T KNOW
 REFUSED

7. Do you currently make a mortgage payment?

- YES
 NO GO TO END

8. What is your mortgage payment each month?

- \$ _____
 DON'T KNOW
 REFUSED

END: Those are all the questions I have for you today. Thank you very much for your participation.

APPENDIX D
TRAINING AGENDA

CPI Housing Tenure Survey Interviewer Training Agenda October 14, 2002

- 9:00 am ID Badges
- 10:00 am Staff Introductions
- 10:15 am Project Overview
- 10:30 am The Case Folder
- 11:30 am The Questionnaires
- 12:15 pm LUNCH and ID Badges**
- 1:00 pm Contact Procedures and Frequently Asked Questions
- 2:00 pm Role Plays (Contact and Questionnaire)
- 3:00 pm Administrative Procedures
-Interviewer Edit

- Mailing Procedures
- Weekly Report Calls
- Time Sheets

4:30 pm Questions and Wrap-up

APPENDIX E
ADVANCE LETTER

U.S. Department of Labor

Bureau of Labor Statistics
2 Massachusetts Ave. N.E.
Washington, D.C. 20212



October, 2002

Dear Sir or Madam:

The Bureau of Labor and Statistics (BLS), a Federal agency in the US Department of Labor, is asking for your help. The BLS is responsible for the production for the Consumer Price Index (CPI), and for that purpose it is trying to improve the way it measures housing costs in neighborhoods like yours. As I am sure you understand, housing expenditures are a major part of everyone's budget and it is important to ensure that the CPI accurately measures how these costs change from year to year.

Your participation in a short survey (approximately 5 minutes) will help us evaluate a new way we are considering to collect this information. We have commissioned Westat, a research company based in Rockville Maryland, to conduct this study.

Your residence has been chosen randomly to be a part of the study. We are asking you to help us by voluntarily participating.

A Westat interviewer will be contacting you to ask a few questions concerning your home. The questions require approximately 5 minutes to complete. Westat and Bureau of Labor Statistics will use the information you provide for statistical purposes only and will hold the information in confidence to the full extent permitted by law.

Sincerely yours,

A handwritten signature in cursive script that reads "John S. Greenlees".

John S. Greenlees
Assistant Commissioner
Consumer Prices and Prices Indexes