APPENDIX F: FINAL SAMPLING PLAN

This plan discusses procedures for selecting the sample and creating the sample weights and adjustments for non-response and undercoverage. This plan was prepared prior to data collection. The final sampling and weighting procedures are fully documented in this codebook.

CONTENTS

TABLES

I. INTRODUCTION

This report presents the sampling plan for the BTS 2000 Omnibus Survey, which evaluates satisfaction with travel on the nation's highways. The sampling plan will be probability based so that study results can be used to make inferences about adults in the U.S. household population. Steps involved in sample design and implementation include: (1) definition of the target population, (2) construction of the sampling frame, (3) specification of sample selection procedures, (4) evaluation of the precision of estimates, and (5) creation of sampling weights and adjustment for nonresponse and undercoverage.

II. THE TARGET POPULATION

The *target population* for a survey is the entire set of population units about which the survey data are to be used to make inferences (Cox & Cohen, 1985). For this survey, the target population is all adults 18 or older in the 50 states and the District of Columbia. We will also constrain the target population to adults in the civilian noninstitutionalized population. To ensure conformity to other national surveys such as the Current Population survey, this population will be defined based upon the definitions of the U.S. Bureau of the Census for the civilian noninstitutionalized population.

III. THE SAMPLING FRAME

The *sampling frame* for a survey is the list or mechanism used to enumerate these population units for sample selection purposes. The sampling frame for this survey will be derived from a list-assisted, random-digit-dialed (RDD) telephone sample approach. Of course, telephone frames exclude those households without telephones, but this source of undercoverage has been steadily declining over time. In 1963, only 80 percent of American households had telephones; by 1988 about 93 percent of all households had telephone service (Thornberry & Massey, 1988). The 1998 Current Population Survey, March Supplement, measured household telephone coverage at 94 percent.

This list-assisted RDD sampling frame provides an innovative solution to the operational problems commonly encountered in the more traditional Mitofsky-Waksberg telephone sampling approach (Waksberg, 1978). Commercial vendors construct these list-based RDD sampling frames by first obtaining a list of all working area code/exchange combinations allocated for residential service (Kulp, 1994). Adding all combinations of digits from 00 to 99 to these six-digit area code/exchange combinations then creates all residential-service hundred-number banks. (These banks are called *hundred-number banks* because they represent the first eight digits of the ten-digit phone number and hence can be linked to 100 unique phone numbers.) In the "list-assisted" step of frame building, all possible hundred-number banks are compared to a frame of listed telephone numbers and the number of residential telephone listings associated with each hundred-number bank is recorded. Finally, geographic coordinates are used to associate location (such as county) and demographic characteristics (such as percent minority) to each hundred-number bank.

We will include in the sampling frame all hundred-number banks that contain at least one listed residential telephone number. Hundred-number banks that have zero residential listings will be

excluded. This exclusion will substantially reduce the incidence of nonworking numbers in the sampling frame. Studies have shown that excluding the zero listed hundred-number banks results in minimal undercoverage bias (Brick, et al., 1995).

IV. SAMPLE SELECTION

For this survey of adults, MPR will develop sample selection procedures that will be used in association with the truncated, list-based frame maintained by Genesys Sampling Systems.[1] The sample will be selected systematically after sorting the frame by the nine Census divisions (New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, and Pacific, which is divided between Alaska and Hawaii and all others) and by urban versus rural counties. Because the counties are divided by division and metropolitan status and carefully ordered, the systematic selection results in a sample that has the equivalent of 20 implicit strata. The underlying sampling frame structure can be conceptualized as a hierarchy. Within each Census division, urban counties are ordered from largest to smallest metropolitan area. Within each metropolitan area, exchanges are ordered by those serving the county containing the central city, followed by those serving the remaining non-central city counties. Within each division, rural counties are ordered in serpentine fashion from north to south and from east to west. The imposition of this implicit stratification will impose geographic representation and reduce the expected sampling variation.

MPR creates RDD samples using an in-house software system developed by Genesys Sampling Systems, which allows us to specify complex sampling designs. The sample is created in-house and as close to the date of interviewing as possible to ensure timeliness. For analysis purposes we will also append a number of variables to the sample file (see Table F-1).

[1]Zero-listed banks are not included in the truncated frame.

TABLE F-1. VARIABLES FROM SAMPLING FRAME.

Income 0K - <10K	Income 0K - <10K%
Income 10K - <15K	Income 10K - <15K%
Income 15K - <25K	Income 15K - <25K%
Income 25K - <35K	Income 25K - <35K%
Income 35K - <50K	Income 35K - <50K%
Income 50K - <75K	Income 50K - <75K%
Income 75K+	Income 75K+%
Age 0 –17	Age 0 -17%
Age 18 – 24	Age 18 - 24%
Age 25 – 34	Age 25 - 34%
Age 35 – 44	Age 35 - 44%
Age 45 – 54	Age 45 - 54%
Age 55 – 64	Age 55 - 64%
Age 65+	Age 65+%
Metropolitan Status Code	Total Number of Households
Primary Zip Code	Percent Owner Occupied
Primary City	Race – Black %
State	Race – Hispanic %
State and County Federal Information	Race – White %
Processing Standards (FIPS) code	Listed Household
Time Zone	

MPR customarily screens all samples for business and nonworking numbers using Genesys' IDS system. First, the file of sampled phone numbers is compared to two business databases and business numbers are removed. Next, the remaining numbers are processed through the dialer to identify nonworking intercept messages. The Genesys Sampling System recently developed an additional screener system, IDplus. The IDplus system starts with the same database comparison to identify known businesses. The remaining numbers are then processed using automated dialing equipment and the phone is allowed to ring up to two times. Specially trained agents are available to speak to anyone who might answer the phone and the number is identified appropriately. This process identifies many more nonresidential numbers and, therefore, could potentially reduce the interviewer time necessary for the survey. For a survey with a short field period, IDplus has the potential to reduce the calendar time needed to process the sample, which could have a beneficial effect on response rates. We propose using IDplus for half the sample and the usual ID operation for the remainder. This embedded experiment will allow us to determine the advantages that IDplus might have for future BTS Omnibus surveys in terms of the time to field the sample. We will also evaluate the consequence of IDplus on response rates. Both A and B questionnaires will have half of their samples prepared using both screening methods. Furthermore, both data collection centers will be administering both instruments so we will be balancing the sample across centers as well. This approach will allow us to determine whether the gains associated with the IDplus methodology warrant future consideration.

To determine the initial sample size, we have to make a number of assumptions. We present below the working residential hit rates and cooperation rates we typically encounter in an RDD telephone survey. However, these assumptions may not be accurate given the limited time frame for this project. Another

unknown factor is the effect of screening half the sample using IDplus, which could lead to a higher percent of working residential numbers. Therefore, we have included a process that will allow us to test these assumptions and adjust the total sample size accordingly. First, an initial sample will be selected based upon optimistic assumptions about response and eligibility rates. This sample will be divided into three replicate samples. Data collected from the first replicate sample will be used to refine our estimate for the response rate and other required assumptions. We can then determine how much additional sample will be necessary to reach the desired number of completed interviews. We anticipate adding a fourth replicate to make up for any short fall associated with less than optimistic response or eligibility rates.

This survey requires that a total of 2,000 completed interviews: 1,000 interviews with the A questionnaire and 1,000 interviews with the B questionnaire. Our previous experience indicates that to achieve this result, 8,045 telephone numbers will need to be screened to determine if the number is a working residential number. From past experience, we expect to be able to determine the residential status for approximately 90 percent of these numbers or 7,241 numbers. Generally, these indeterminable residency-status numbers are "ring no answer" cases (after 8 callbacks) and tend to be unassigned numbers. Again, our experience suggests that roughly 50 percent of these 7,241 numbers will be identified as residential numbers for a total of 3,620 identified residential numbers. Having identified a number as residential, we will need to "roster" adult household members. We anticipate that 90 percent of the identified residential numbers will provide the roster information or 3,258 households. We anticipate that of the 3,258 households completing the roster 99 percent will have an eligible adult or 3,226 households. Having completed the roster, we will then randomly select an adult from the list and randomly assign them to subgroup A or subgroup B. From the 3,226 eligible sampled adults, we expect that 62 percent will cooperate with the interview to yield the required 2,000 interviews, that is, 1,000 completed interviews for subgroup A and 1,000 completed interviews for subgroup B.

V. PRECISION

Let us turn now to the precision anticipated under the proposed RDD design. To assess the efficiency of estimated percentages $\hat{P}$, it is useful to examine the half-length of confidence intervals around the estimate. For this application, the confidence interval can be approximated for design purposes as:

$$\hat{P} \pm z_{1-a} \sqrt{Var(\hat{P})}.$$

Here z_{1-a} is value of the critical point x at which the normal cumulative distribution function equals 1-% (i.e., $F(x)$=%). The half-length HL is:

$$HL = z_{1-a} \sqrt{Var(\hat{P})}.$$

That is, $\hat{P}$ can be expected to fall within the range [P-HL, P+HL] with 95 percent confidence for the proposed sample sizes. Therefore with a sample size of 2,000 and P=50 percent, the confidence

interval range would be [50 – 2.19, 50 + 2.19]. Table F-2 presents the half-length interval for domains of various sizes. A domain is a subgroup for which separate analyses will be made.

TABLE F-2. THE HALF LENGTH OF 95% CONFIDENCE INTERVALS IN PERCENTAGE POINTS FOR VARIOUS PERCENTAGES BEING ESTIMATED FOR DOMAINS OF VARIOUS SIZES.

Percentage	***Domain Sample Size***								
P	100	200	300	400	500	1000	1500	2000	3000
5	4.27	3.02	2.47	2.14	1.91	1.35	1.10	0.96	0.78
10	5.88	4.16	3.39	2.94	2.63	1.86	1.52	1.31	1.07
15	7.00	4.95	4.04	3.50	3.13	2.21	1.81	1.56	1.28
20	7.84	5.54	4.53	3.92	3.51	2.48	2.02	1.75	1.43
25	8.49	6.00	4.90	4.24	3.80	2.68	2.19	1.90	1.55
30	8.98	6.35	5.19	4.49	4.02	2.84	2.32	2.01	1.64
40	9.60	6.79	5.54	4.80	4.29	3.04	2.48	2.15	1.75
50	9.80	6.93	5.66	4.90	4.38	3.10	2.53	2.19	1.79
60	9.60	6.79	5.54	4.80	4.29	3.04	2.48	2.15	1.75
70	8.98	6.35	5.19	4.49	4.02	2.84	2.32	2.01	1.64
80	7.84	5.54	4.53	3.92	3.51	2.48	2.02	1.75	1.43
90	5.88	4.16	3.39	2.94	2.63	1.86	1.52	1.31	1.07
95	4.27	3.02	2.47	2.14	1.91	1.35	1.10	0.96	0.78

VI. SAMPLING WEIGHTS AND ADJUSTMENTS

Probabilities of selection will be computed and maintained for each level of sampling. When data collection is complete, sampling weights will be calculated as the inverse of the probability of selection. The sampling weights will be adjusted to compensate for nonresponse at each stage of the selection and interviewing process. The extent to which members of the target population are missing from the survey frame is referred to as *undercoverage*. Poststratification adjustments will be made to the nonresponse-adjusted weights to compensate for undercoverage of nontelephone households. We will poststratify to the cross-classification of age, race, and sex for all adults using population projections derived from the 1990 Decennial Census and extrapolated to the present. For further information about the weighting process, see Cox and Cohen (1985, Chapter 7) or Cox, 1991. MPR will calculate weights for the total sample and each half sample, that is subgroup A and subgroup B.

REFERENCES

Brick, J. Michael, Waksberg, Joseph, Kulp, Dale, and Starer, Amy (1995). "Bias in List-Assisted Telephone Samples". *Public Opinion Quarterly*. Vol 59: 218-235.

Cox, Brenda G., and Cohen, Steven B. (1985). *Methodological Issues for Health Care Surveys*, New York: Marcel Dekker Inc.

Cox, Brenda G. (1991). "Weighting Survey Data for Analysis," American Statistical Association Continuing Education Program Tutorial.

Kulp, Dale W. (1994). "Dynamics of 'List Assisted' Random Digit Dialing (RDD) Frame Coverage," *Proceedings of the American Statistical Association, Survey Research Methods Section*.

Thornberry, Owen T., Jr., and Massey, James T. (1988). "Trends in United States Telephone Coverage Across Time and Subgroups," in R. M. Groves, P.. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls, and J. Waksberg (eds.), *Telephone Survey Methodology*, New York: John Wiley & Sons, pp. 25-50.

Waksberg, J. (1978). "Sampling Methods for Random Digit Dialing," *Journal of the American Statistical Association*, 73, 40-46.