Alan Sunter
63 Fifth Avenue
Ottawa, Ontario K1S 2M3

January 15, 1999

Mr. Michel Rochon
151 rue Jolicoeur
Hull, Quebec
18Z 1C8

Dear Michel:

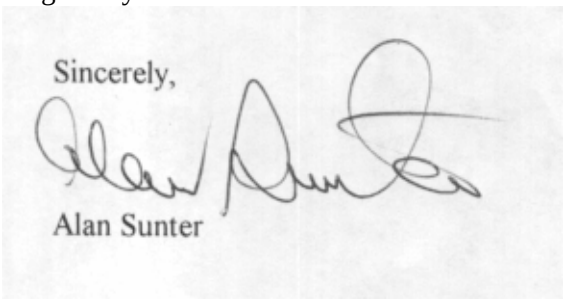
re: ASDE Survey Sampler

I have reviewed the changes in the ASDE Survey Sampler that you have made since my earlier evaluations of August 2, 1995 and March 7, 1996 and am satisfied that the remarks made in those evaluations continue to reflect my views on your methodology.

The changes may generally be described as operational upgrades. They considerably enhance the utility of the system without, however, detracting from its original statistical integrity.

I would be pleased to recommend your software to any firm with a requirement for RDD and/or listed phone number sampling.

Original by:



Annex A Canada Survey Sampler
Sunter Evaluations

Mr. Michel Rochon
151 rue Jolicoeur
Hull, Quebec
J8Z lC8

August 2 1995
Revised February 21 1999

Dear Michel:

Following your request for a disinterested technical evaluation of the ASDE Survey Sampler (CSS) procedures for sampling. I have examined in detail both your customer documentation and the documented audit trail for the selection of a particular stratified sample for the whole of Canada. This is in addition to the demonstration of the software and its built-in options that you provided earlier.

I can confirm that the procedure as implemented is consistent with the customer documentation and, further that the customer documentation accurately reflects the merits and potential uses of the software.

With regard to the technical characteristics of the sampling procedures my analysis demonstrates that both for listed and unlisted numbers the samples drawn by the CSS procedures select members of the respective populations with equal probability and without replacement within each stratum. The listed number sample is of fixed size, as determined by the sample allocation option chosen by the user; the unlisted number sample has some sampling variance in the sample size actually realised but this variance (which appears, in any event to be unavoidable) has negligible impact in my view on the efficiency of the overall design. The sampling variances of stratum estimates will be equivalent to or less than those obtained from simple random samples of equivalent sizes. Finally, I agree that "validity likelihood" factors based on the ratio of listed numbers to all numbers within a block is a reasonable a priori estimate of the ratio of valid unlisted numbers to all unlisted numbers within a block and that it should be useful in a practical survey application.

I congratulate you on a successful solution to the technical problems of sampling to meet RDD requirements while maintaining the advantages of listed phone samples from electronic databases and on the development of the ASDE Survey Sampler. I have no hesitation in recommending your software to any firm with a requirement for RDD or/and listed general population phone sampling.

Sincerely,

Alan Sunter, President

A.B. Sunter tel (613) 235-4172 fax (613) 235-0154

7 March, 1996
Revised February 21 1999

Mr. Michel Rochon
151 rue Jolicoeur
Hull, Qu6bee
J8Z IC8

Dear Michel:

You have asked me to compare your sampling procedure with the more usual Random Digit Dialling (RDD) procedure often specified as "required" in government "requests for proposal".

Your procedure, as I have demonstrated in an earlier analysis, selects both listed and unlisted numbers with equal probability and without replacement within each stratum (usually an exchange or a block of exchanges). The listed number sample is of fixed size, as determined by the sample allocation option chosen by the user; the unlisted number sample has some sampling variance in the sample size actually realised but this variance (which appears, in any event, to be unavoidable) has negligible impact in my view on the efficiency of the overall design, The sampling variances of stratum estimates will be equivalent to, or less than, those obtained from simple random samples of equivalent sizes. I also suggested at that time that a "validity likelihood" factor based on the ratio of listed numbers to all numbers within a block would be a reasonable a priori estimate of the ratio of valid unlisted numbers to all unlisted numbers within a block and that it should be useful in a practical survey application.

In order to compare your procedure with the usual procedure for RDD, it will be useful first to provide a little detail on how it actually works. You first specify the size of a "block" of numbers (usually but not necessarily a nominal 1000, equivalent to the set of four digit suffixes to a three-digit exchange). You then select a sample of blocks with <u>probability proportional to the number of listed numbers in each block.</u> Random selections of numbers within the block are then made until a listed residential number is selected. Intervening unlisted numbers are retained and form the equal probability sample of such numbers.

A.B. Sunter tel (613) 235-4172 fax (613) 235-0154

Now, the usual procedure for RDD is still the procedure described in "Sampling Methods for Random Digit Dialling", Joseph Waksberg, *J. American Stat. Assoc., 73, 40-46.* A version of this procedure, comparable to that just described for your procedure, would form the same blocks of numbers but would then select them depending on whether a single random number selection within the block reached a valid number.

It is easy to see that this procedure selects blocks with probability proportional to the number of valid numbers in each block. Once a block is selected, the design may call for a certain fixed number of valid calls to be achieved within the block. Similarly, your own method may call for a certain number of listed numbers to be selected and called, with intervening unlisted numbers also to be selected and called.

Both methods yield equal probability samples of valid numbers and, in terms of their relative statistical efficiency for unbiased estimates (all selected numbers to be called and questions of non-response aside), there is really nothing to choose between them. In practical terms, however, I think that your procedure bas the edge. This is because you know in advance which of your selected numbers are listed numbers (with relatively high likelihood of being a productive call) and which are unlisted (with relatively low likelihood), and a little experience with actual surveys should enable the user to predict the likelihood fairly accurately. Supposing, that experience shows, as I think it will, that the number of valid unlisted numbers in a block tends to be proportional to the number of listed numbers and hence that the likelihood of an unlisted number being valid is proportional to L_h/N_h,, where L_h is the number of listed numbers in the h-th block and N_h, total numbers, (The constant of proportionality may be of the order of a few percent, but can only be established by experience over one or more surveys.) A reasonable procedure would be to order selected unlisted numbers by their likelihood of being valid, then beginning at the top of the list continue to make calls until the "remaining likelihood" is less than (say, one half of) the non-response incurred in calls to valid numbers so far (assuming that all selected listed numbers were called first.) For example, if 1000 calls to valid numbers have had non-response of, say, 100 then we might stop calling the listed numbers when the remaining likelihood is less than 50.

Consider the following comparison of the two procedures. Suppose that we have 80 blocks of 100 sequential telephone numbers, the first 10 of which have 90 listed numbers each, the second 10 each have 80 listed numbers, the third 10 each have 70 listed numbers, and so on down to the last 10 which each have 20 listed numbers. In each block, let us suppose, 90% of the **listed numbers** are valid, and that it is known from previous experience that the number of valid **unlisted numbers** is about 4% of the listed numbers. Now suppose we use the Waksberg RDD procedure to select blocks with probability proportional to number of valid numbers and use the number, which selected the block as its single sample unit for that block.

Alternatively we can use your procedure to select a comparable number of blocks and then select numbers within that block until exactly one listed number has been selected. It is easily demonstrated (see Table) that the Waksberg procedure, which will make one call within each of the 80 blocks, has an expected yield of 41.4 calls to valid numbers (for 80 calls overall). Now suppose that you use your procedure to select, say 45 blocks with probability proportional to number of listed numbers, before proceeding further as described above. It is now easy to demonstrate that you will select exactly 45 listed numbers together with an expected 36.8 unlisted numbers which will yield an expected 1.8 valid numbers. Of your 45 listed numbers an expected 40.5 will be valid, giving you a total of 81.8 calls for an expected 42.3 valid numbers, as compared with the RDD 80 calls for an expected 41.4 valid numbers. In these terms there is, as I said earlier, little to choose between the two procedures. However, suppose your initial 40.5 calls to valid listed numbers produced, say, 5 non-responses. In this case there would be little justification for making a further (expected) 36.8 calls in order to obtain a further (expected) 1.8 valid numbers, The essential difference between the two procedures, from an operational point of view, is that your procedure gives you an opportunity to make this kind of informed judgement while the RDD procedure does not.

Sincerely,

Alan Sunter, President

A.B. Sunter tel (613) 235-4172 fax (613) 235-0154

Annex A Canada Survey Sampler
Sunter Evaluations

Comparison of two sampling procedures

Block	Listed		Unlisted		Block Selection Probability		Cansam Expected Yield		
	Valid	Invalid	Valid	Invalid	RDD	Cansam	Listed	Unlisted	Unlist.Valid
1	81	9	3.6	6.4	0.85	0.92	0.92	0.10	0.0368
...	...	...	...	...	...	...	...	...	...
10	81	9	3.6	6.4	0.85	0.92	0.92	0.10	0.0368
11	72	8	3.2	16.8	0.76	0.82	0.82	0.20	0.0327
...	...	...	...	...	...	...	...	...	...
20	72	8	3.2	16.8	0.76	0.82	0.82	0.20	0.0327
21	63	7	2.8	27.2	0.66	0.72	0.72	0.31	0.0286
...	...	...	...	...	...	...	...	...	...
30	63	7	2.8	27.2	0.66	0.72	0.72	0.31	0.0286
31	54	6	2.4	37.6	0.54	0.61	0.61	0.41	0.0245
...	...	...	...	...	...	...	...	...	...
40	54	6	2.4	37.6	0.54	0.61	0.61	0.41	0.0245
41	45	5	2	48	0.47	0.51	0.51	0.61	0.0205
...	...	...	...	...	...	...	...	...	...
50	45	5	2	48	0.47	0.51	0.51	0.51	0.0205
51	36	4	1.6	58.4	0.37	0.41	0.41	0.61	0.0164
...	...	...	...	...	...	...	...	...	...
60	36	4	1.6	58.4	0.37	0.41	0.41	0.61	0.0164
61	27	3	1.2	68.8	0.28	0.31	0.31	0.72	0.0123
...	...	...	...	...	...	...	...	...	...
70	27	3	1.2	68.8	0.28	0.31	0.31	0.72	0.0123
71	18	2	0.8	79.2	0.19	0.20	0.20	0.82	0.0082
...	...	...	...	...	...	...	...	...	...
80	18	2	0.8	79.2	0.19	0.20	0.20	0.82	0.0082
Total	3960	440	220	3380	41.24	45.00	45.00	36.82	1.8000

A.B. Sunter tel (613) 235-4172 fax (613) 235-0154

Annex A Canada Survey Sampler
Sunter Evaluations

Mr. Michel Rochon
151 rue Jolicoeur
Hull, Quebec
J8Z 1C8

April 26, 2001

Dear Michel:

re: ASDE Survey Sampler

After reviewing your system, as upgraded since my last review, I continue to be very favorably impressed by its operational convenience and, perhaps more importantly, by its potential for statistical efficiency relative to the "classical" approach of Random Digit Dialling (RDD) of numbers whose listed/unlisted status is unknown.

As in my 1995 review, I can confirm that the system selects members of the listed/unlisted categories in each stratum with equal probability and without replacement and that the sampling variances of stratum estimates will be equivalent to, or less than, those obtained from simple random samples of equivalent size.

The operational convenience of the system is best appreciated, of course, by a demonstration or, better, a "hands on" trial. My interest here is with the potential for statistical efficiency; i.e. with the potential for maximizing the precision of estimates from survey samples at a specified cost or, alternatively, for minimizing cost at a specified precision.

The first element of such efficiency is the potential that your system gives for precise geographic targeting and/or stratification. By "targeting" I refer to your capability of minimizing the inclusion of telephone subscribers who do not fall into the population of interest. By "stratification" I refer to your ability for independent and controlled sampling for separate specified geographic domains. This ability allows the user to allocate his sample to strata according to the allocation algorithm that most efficiently meets the survey objectives.

The second element of potential efficiency results from your precise information as to the numbers of listed and unlisted numbers in each stratum, and hence, in your ability to estimate in advance the number of in-service numbers in each of these categories. With this information, and with a background of experience in telephone surveys that your subscribers will have, it is possible to make quite precise estimates in advance of the unit costs of responses in each of the two categories and, hence, to optimize sample allocation to the listed/unlisted categories (or sub-strata) of each stratum. In the typical situation of unit costs of responses from unlisted numbers many times that for listed numbers, my simulations suggest that optimal allocation will typically lead to reductions in the

number of calls to unlisted numbers of 50% or more, with a compensating increase of about 10% in calls to listed numbers (so as to keep the costs constant) and a significant increase in precision. Alternatively, costs may be reduced by 10% while keeping precision constant. There is always a gain[1] in efficiency from optimal allocation; it is only the magnitude of the gain that will vary.

I congratulate you on an attractively packaged and priced product which provides your subscribers with the potential for significant gains in statistical and operational efficiency.

Sincerely,

Alan Sunter

[1]This assumes, of course, that the estimation procedures will take proper account of differences in sampling rates and also of differences in response rates.

A.B. Sunter tel (613) 235-4172 fax (613) 235-0154