

Incentive Effect on Response Rates for the 1997 Survey of Household Spending

Ben Vandermeer, Statistics Canada

RH Coats Bldg, 16th Floor, Tunney's Pasture, Ottawa ON, K1A 0T6

Key Words: Incentives, Response Rate, Assignment Effect, Respondent Relations

A. INTRODUCTION

Prior to 1997, the Family Expenditures Survey (FAMEX) was conducted, at the national level, every four years to gather very detailed expenditure information for a given calendar year. The sample, which was selected from the Labour Force Survey sampling frame, usually contained around 14,000 households with an average interview length of about 2.5 hours. The final response rate was only 77% in 1996, despite it being a mandatory survey.

Due to the initiative of the Project to Improve Provincial Economic Statistics (PIPES), starting with the 1997 calendar year FAMEX underwent a major overhaul. It is now known as the Survey of Household Spending (SHS) and is conducted every year with an increased sample of about 23,000 households. In addition, many survey questions were collapsed in an attempt to shorten the length of the survey and decrease response burden. Finally, it was decided to try incentives as a method of improving response rates.

While Statistics Canada does not normally offer incentives to respondents for its surveys, it was felt the Survey of Household Spending may be one survey that was worth a test. The 1992 FAMEX survey had an all-time low response rate of 73.8%. In an attempt to combat this, the 1996 survey was made mandatory. While this did increase the response rate (77.4%), it resulted in public controversy. With a new name and format, the 1997 survey was returned to voluntary status for 1997 and a test of incentives was performed to try to improve respondent relations as well as increase the survey response rate. This document will outline the experimental design and the final analysis of the results.

B. EXPERIMENTAL DESIGN

1. Incentive Allocation

Two different incentives were used in the experiment. The first was a one-year subscription to the Statistics Canada publication *Canadian Social Trends*. The second was a telephone calling card good for twenty

minutes of long distance calling anywhere in North America. These incentives were offered at the beginning of the interview and it was made clear to the respondent that their receiving of the gift would be contingent upon their responding to the questionnaire.

Every dwelling was assigned to one of the three experimental groups in a split of 40% to control and 30% to each of the two incentives. Also, for operational reasons, an interviewer worked with only one of the two incentives; that is to say one interviewer would have some dwellings in the control group and some dwellings in one (and only one) of the incentive groups. Finally to avoid any possible negative reaction to the study (i.e. "my neighbor got a gift for answering your survey and I didn't") every dwelling within a cluster was placed in the same experimental group.

To achieve these goals, a list of all the interviewer assignments with listings of all the clusters in each assignment was obtained. Each assignment was randomly designated as either magazine (M) or telephone card (T) ensuring an even split within each region (Atlantic, Quebec, Ontario, Prairies, and BC). Then for every interviewer assignment each cluster was randomly designated as either control or incentive making the split as close to 40%-60% as was possible with the number of clusters in the assignment. This would give us a split with an expected value of 40%-30%-30% for control-magazine-telephone card at the dwelling level—the actual split obtained was about 41%-30%-29%.

2. Response Definitions

The purpose of this test was to determine whether the incentive had a significant effect on response rates. Normally when an interviewer is unable to contact a respondent or the interview is prevented due to unusual circumstances, the household in question will be counted as a non-response. In these instances the respondent was never given the opportunity to participate or refuse to participate in the survey and thus was never (for those in an incentive group) informed about the incentive they would receive for their participation. Since these households give us no information about the incentive's effect on willingness to respond, they were excluded from our target population. Only dwellings where some contact was made were considered for this study.

Table 1: Canada Level Response Rates

<u>Incentive</u>	<u>Response</u>	<u>Non Response</u>	<u>Total</u>	<u>Response Rate</u>	<u>P-Value</u>
Control	7464	1658	9122	81.8	
Magazine	5630	1151	6781	83.0	0.026
Phone Card	5397	1164	6561	82.3	0.249
Total	18491	3973	22464	82.3	

C. ANALYSIS

1. Analysis of raw data at various levels

The raw data that came out of the experiment was analyzed by looking at the response rates for each of the three incentive groups as well as comparisons between each of the incentive groups with the control group.

The statistical test used in these comparisons was the one sided Fisher's exact test for 2 × 2 contingency tables. By using a one-sided test we are making the assumption the incentive does not have a harmful effect on the response rate; i.e. a respondent is not less likely to respond when offered an incentive.

We start by looking at the data at the national level (see Table 1). We see that of the three groups, the magazine had the highest response rate at 83.0%. This was significantly different from the control group's rate of 81.8%. While the telephone card's response rate was also better than the control group's, it was not significantly so.

Many confounding factors could account for the difference shown in Table 1. Since it is not unreasonable to assume that respondents could react differently to incentives in different regions, response rates were computed for each region.

Splitting the respondents into the five main regions of Canada (Atlantic, Quebec, Ontario, Prairie, and B.C) showed that the response rate for the magazine is better than that of the control group in all regions except Quebec. It is quite significantly better in the Atlantic region (p-value < 0.01) but the difference is not significant in any other region.

For the phone card the response rates were superior to the control group in all regions except the Atlantic. None of the differences are significant.

All these results were somewhat puzzling based on other information that was available. The results of a separate respondent relations study showed that both respondents and interviewers were having much more favorable reactions to the telephone card than they were to the magazine. Also, the respondents receiving the magazine subscription had to mail back a subscription form to Ottawa head office, and the counts of the number of subscription requests received from SHS magazine respondents showed that only about 22% even bothered to subscribe. The results obtained in these tests seemed somewhat counter-intuitive to these other findings.

2. Assignment Grouping Tests

All the p-values obtained so far using Fisher's exact test are good in the sense that they use all the data collected to its full potential use, but weaker in the sense that they require assumptions about that data that is not easy to verify. These include assumptions that there are no confounding factors affecting our results. One of these potential factors is an interviewer effect.

One possible method of measuring an interviewer effect is through experience of the interviewer, as was done in the previous study of incentives for the 1990 FAMEX survey. This time around the only information available was concerning which interviewers had experience on the Labour Force Survey. Our analyses showed that this factor did not seem to have any effect on response rates (as was the case for the previous study). This of course does not eliminate the possibility

Table 2: Assignment Grouping Tests at Canada Level

<u>Incentive</u>	<u>Positive Difference</u>	<u>Negative Difference</u>	<u>Total</u>	<u>P-Values</u>		
				<u>Sign Test</u>	<u>Mann-Whitney</u>	<u>Paired-t</u>
Magazine	177	181	358	0.604	0.663	0.699
Phone Card	188	163	351	0.100	0.177	0.180
Total	365	344	709	0.226	0.361	0.382

of an interviewer effect, since some interviewers will always elicit a greater response rate than others regardless of experience.

The design of the experiment allows us a unique opportunity to completely remove the potential interviewer effect from our test. Since every interviewer had households in the control group and in one and only one of the incentive groups, each interviewer assignment can be isolated to determine whether the incentive respondents or the control respondents had a better response rate. By doing this for all interviewers we can make statistical tests that cannot be confounded by an interviewer factor since only households within one interviewer’s assignment are being compared.

Under a null hypothesis that the incentive has no effect on response rates it would be expected that an equal number of interviewers would have a better response rate in their control households than in their incentive households. The sign test, Mann-Whitney test, and the paired t-test were performed on these differences. Table 2 above gives the results of these three tests performed on the entire sample of interviewer assignments. The

heading *Positive Difference* indicates the number of interviewer assignments where the incentive response rate was greater than the control response rate, while *Negative Difference* indicates the opposite.

The results seen here are quite different from those observed in the previous tests. For the magazine, there were actually slightly more assignments that had better control response rates. This would indicate no effect for the magazine at the Canada level—quite a bit different from the significance indicated in the raw data tests. For the telephone card there were 188 positive differences opposed to 163 negative differences, giving a sign-test p-value of 0.10—close to significant. The Mann-Whitney and Paired-t tests performed on the ranks and magnitudes of the proportions are higher—less significance is shown when the values (and not just the signs) are taken into account.

All interviewer assignments were confined to one region; thus we can regionally break up Table 2 above and check the results. Table 3 below shows the results for the Atlantic and B.C. regions (the Quebec, Ontario, and Prairie Regions had no significant results).

Table 3: Assignment Grouping Tests at Regional Level

<u>Region</u>	<u>Incentive</u>	<u>Positive Difference</u>	<u>Negative Difference</u>	<u>Total</u>	<u>P-Values</u>		
					<u>Sign Test</u>	<u>Mann-Whitney</u>	<u>Paired-t</u>
ATLANTIC	Magazine	48	52	100	0.691	0.634	0.698
	Phone Card	53	43	96	0.179	0.114	0.116
	Total	101	95	196	0.361	0.267	0.337
BC	Magazine	25	30	55	0.791	0.688	0.725
	Phone Card	33	24	57	0.145	0.102	0.081
	Total	58	54	112	0.388	0.289	0.236

These results seem very odd compared to the previous results obtained from the Fisher's exact tests done in the previous section. Recall from those tests that we observed a significant difference in response rates for the magazine in the Atlantic region, but no significant results in any other region for either incentive. The story is quite different in these tests, with the telephone card performing much better than the magazine in the comparisons. The telephone card has consistently low p-values (although not quite at significance level) in the Atlantic and BC regions. The magazine has nothing close to significance anywhere and, most surprisingly of all, fairs quite poorly in the Atlantic region where previous tests had indicated it to be highly significant.

What is the cause of these great differences in the results of our tests? Why does the same data seem so much different when it is looked at it in another way? These are certainly questions that need to be answered. The key here is the assumptions being made. The Fisher's tests performed in the previous section assume that the control group is the same as each incentive group in every way that would effect the resulting response rate with the exception of the main incentive factor that is being tested. Considering that the incentives were allocated randomly, these assumptions seemed reasonable; however, one thing that the allocation did not take into account was the interviewers. The interviewers were placed in their

assignments by the regional offices *after* the incentive allocation was completed.

The assignment grouping tests performed here are controlling for any interviewer effect by comparing response rates within an assignment. Even if the assumption that the interviewers assigned to the magazine group elicit the same response rate as those assigned to the telephone card group is not valid these less powerful assignment grouping tests will still be valid. In the next section this assumption is tested.

3. Control Split Testing

The Fisher's Exact Tests performed in the first part of this section will be done again here with a slight twist to try to control for potential interviewer bias. Rather than comparing the entire control group with each of the two incentive groups (as was done previously) we will instead compare each incentive group with the control group that was handled by the same set of interviewers. That is to say, all respondents who received the magazine incentive will be compared with only those respondents in the control group who were interviewed by the same group of interviewers. At the same time the two control groups will also be compared to see if there is evidence of difference between the 2 groups. All p-values given in this section are again using the one-sided Fisher's Exact Test.

Table 4: Control Split Tests at Canada Level

<u>Magazine</u>						
<u>Non</u>						
<u>Incentive</u>	<u>Response</u>	<u>Response</u>	<u>Total</u>	<u>Response Rate</u>	<u>P-Value</u>	
Control	3734	776	4510	82.8		
Magazine	5638	1151	6789	83.0	0.373	
Total	9372	1927	11299	82.9		
<u>Phone Card</u>						
<u>Non</u>						
<u>Incentive</u>	<u>Response</u>	<u>Response</u>	<u>Total</u>	<u>Response Rate</u>	<u>P-Value</u>	
Control	3583	866	4449	80.5		
Phone Card	5399	1164	6563	82.3	0.012	
Total	8982	2030	11012	81.6		

Comparisons between magazine and phone card assignments

2 Control Groups: p-value = **0.003**

2 Incentive Groups: p-value = 0.121

The results are quite different when the control group is split. In Table 4 the telephone card shows a significant difference in response rate compared to its control group. The magazine's response rate was not significantly different from its control. The reason for this reversal of our original tests can be seen by looking at the response rates within the table and in the comparisons between magazine and phone card assignments shown at the bottom of Table 4.

These two p-values are based upon a test that the magazine group has a superior response rate to the telephone card group. The first p-value compares the 2 control groups—i.e. the magazine control group had a response rate of 82.8% while the telephone card control group had a response rate of 80.5%. The value of 0.003 is highly significant and implies the difference in response rate is not due to chance alone. But it is also not due to an incentive since only those households that did *not* receive an incentive are being examined. In fact, when we do the comparison among households that did receive the incentive (comparing the magazine's 83.0% rate with the telephone card's 82.3% rate) the difference is somewhat less significant with a p-value of 0.121. This is an indication of the much greater increase in response rate for the telephone card group than for the magazine group.

We next broke down Table 4 and presented it for each of the 5 regions to see whether the observed difference between magazine assignments and telephone card assignments was specific to any region. A summary of those tables is shown in Table 5.

For the Quebec, Ontario, and Prairie regions there is no evidence that the magazine interviewers got better response rates than their telephone card counterparts. The p-values obtained comparing the control and incentive groups do not differ greatly from those observed previously when the control group was not split. Once again we do not see any significant p-values.

The story is quite different for the Atlantic and BC regions, as seen on table 5. Both regions show evidence of a difference between the split control groups—indicating a probable interviewer effect on our previous results. The Atlantic region shows an overwhelming difference between the two groups with the magazine interviewers getting a response rate nearly 5 percentage points better than their telephone card counterparts—in the control group—giving a p-value less than 0.0005. Within the groups, the magazine did not significantly increase response rate, while the telephone card did, with a p-value just over 0.05 (0.057)—this reverses our earlier conclusions. A similar situation exists in the BC region; once again there is a significant difference between the two interviewer groups. Comparing within these groups we see the magazine did not result in a significant increase in response rate (in fact it went down!) while the telephone card had an increased response rate over its control group that resulted in a p-value of 0.06—just above the 5% mark.

Table 5: Control Split Tests at Regional Level

Region	Incentive	Group	Response Rate	P-Value of Incentive Comparison	P-Value of Two Control Group Comparison
Atlantic	Magazine	Control	85.9	0.408	0.000
		Incentive	86.2		
BC	Magazine	Control	81.1	0.057	0.013
		Incentive	83.3		
BC	Telephone Card	Control	78.5	0.852	0.013
		Incentive	76.4		
BC	Telephone Card	Control	73.1	0.060	0.013
		Incentive	76.6		

D. CONCLUSIONS

Many tests and analyses have been performed on the response rate data. We can safely eliminate the preliminary set of tests, where the raw data was used, as misleading. These tests do not control for the interviewer effect we have seen (from the third set of tests) exists and thus any increases in response rate observed are not necessarily due to an incentive effect.

The second set of tests completely controlled for the interviewer effect, but this came at a price. By taking each interviewer assignment as one observation our sample size was reduced from the over 22,000 households to the just over 700 interviewer assignments. This reduced sample makes it less likely that an existing incentive effect would be detected. Although there were no significant p-values in any of the tests performed here, there were some indications that the telephone card had some effect particularly in the Atlantic and BC regions where the p-values for the Mann-Whitney and paired t-tests hovered around 0.10. It was clear that the telephone card seemed to perform better than the magazine at the overall level as well as in the Atlantic region—quite the opposite of what the analysis of the raw data told us. This indicated that there was an interviewer assignment effect that could not be ignored.

The Fisher's exact tests done on the raw data were redone by comparing each incentive group with only the portion of the control group that was handled by the same set of interviewers, thus taking this assignment effect into account. In this third and final set of tests the hypothesis of an interviewer effect, suggested by the differences between the first two set of tests, was confirmed. In two regions, the Atlantic and BC regions, there was strong evidence that the group of interviewers that handled the magazine incentive elicited a better response rate than those that handled the telephone card—a difference that was not attributable to the different incentives. The effect in these two regions caused an overall effect at the national level. As stated before this invalidated the results of the original tests that didn't separate the control groups for the two groups of interviewers.

The magazine group had an increased overall response rate of only 0.2 percentage points (from 82.8% to 83.0%), providing no evidence that it has a significant impact on overall response rates. Regionally, it had showed a positive increase in the Atlantic, Ontario, and Prairie regions and a negative increase in the Quebec and BC regions. None of these effects were significant,

and we can conclude that the magazine did not seem to have any effect on response rates.

At the national level, the telephone card showed an increase of almost 2 percentage points in response rate (80.5% to 82.3%), providing significant evidence that it has a positive impact on the overall response rate. Breaking the data down regionally shows that the telephone card had a positive increase in every region with the increases in the Atlantic and BC regions having p-values just above the 5% level. Thus the overall increase in response rate due to the telephone card can be largely attributed to its effect in these two regions.

Although useable information was salvaged from the experiment, the misleading results in the initial tests suggest that perhaps we should not have each interviewer working with only one incentive, even if it is operationally convenient, since it has the potential to jeopardize the experiment.

E. ACKNOWLEDGEMENTS

I would like to thank Mamta Singh and Marc St-Denis for doing the analysis of the respondent and interviewer feedback questionnaires, which fit in very nicely with this study. I would also like to thank Johanne Tremblay and Sophie Arsenault for their many helpful comments over the course of the work. Finally, I would like to thank Éric Rancourt, MP Singh, Jack Gambino, and Maryanne Webber who also took the time to read a first draft of this paper and give their comments.

F. REFERENCES

1. Agresti, Alan (1990): *Categorical Data Analysis*. New York: Wiley.
2. Kumar, S. and Durning, A. (1992): *The Impact of Incentives on the Response Rates for FAMEX 1990: An Evaluation*. Statistics Canada, Internal Document, Methodology Branch.
3. Lehmann, E.L. (1975): *Nonparametrics: Statistical Methods Based on Ranks*. San Francisco: Holden-Day, Inc.
4. Singh, Mamta and St-Denis, Marc (1998): *What Do They Really Think? Incentives and Their Qualitative Impact on Respondents and Interviewers*. Statistics Canada, Internal Document.