



Income Statistics Division

62F0026MIE - 00004

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

Prepared by:
Ben Vandermeer

August 2000



Statistics
Canada

Statistique
Canada

Canada

Data in many forms

Statistics Canada disseminates data in a variety of forms. In addition to publications, both standard and special tabulations are offered. Data are available on the Internet, compact disc, diskette, computer printouts, microfiche and microfilm, and magnetic tape. Maps and other geographic reference materials are available for some types of data. Direct online access to aggregated information is possible through CANSIM, Statistics Canada's machine-readable database and retrieval system.

How to obtain more information

Inquiries about this product and related statistics or services should be directed to: Client Services Income Statistics Division Canada, Ottawa, Ontario, K1A 0T6 ((613) 951-7355; (888) 297-7355; income@statcan.ca) or to the Statistics Canada Regional Reference Centre in:

Halifax	(902) 426-5331	Regina	(306) 780-5405
Montréal	(514) 283-5725	Edmonton	(403) 495-3027
Ottawa	(613) 951-8116	Calgary	(403) 292-6717
Toronto	(416) 973-6586	Vancouver	(604) 666-3691
Winnipeg	(204) 983-4020		

You can also visit our World Wide Web site: <http://www.statcan.ca>

Toll-free access is provided **for all users who reside outside the local dialing area** of any of the Regional Reference Centres.

National enquiries line	1 800 263-1136
National telecommunications device for the hearing impaired	1 800 363-7629
Order-only line (Canada and United States)	1 800 267-6677

Ordering/Subscription information

All prices exclude sales tax

Catalogue no. 62F0026MIE-00004, is available on internet for CDN \$0.00 per issue. Users can obtain single issues at <http://www.statcan.ca/cgi-bin/downpub/freepub.cgi>.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the agency has developed standards of service which its employees observe in serving its clients. To obtain a copy of these service standards, please contact your nearest Statistics Canada Regional Reference Centre.



Statistics Canada
Income Statistics Division

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

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2000

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission from Licence Services, Marketing Division, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6.

August 2000

Catalogue no. 62F0026MIE - 00004
ISSN 0000-0000

Catalogue no. 62F0026MPE - 00004
ISSN 0000-0000

Frequency: Irr.

Ottawa

La version française de cette publication est disponible sur demande

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

ELECTRONIC PUBLICATIONS AVAILABLE AT
www.statcan.ca



TABLE OF CONTENTS

A. EXECUTIVE SUMMARY.....	7
B. INTRODUCTION	9
C. BACKGROUND	9
1. Why Incentives?.....	9
2. Tests of Incentives for Other Surveys	10
3. Incentive Test Performed for FAMEX 1990	11
D. EXPERIMENTAL DESIGN	12
1. Incentive Allocation	12
2. Response Definitions.....	12
E. ANALYSIS	13
1. Analysis of raw data at various levels	13
2. Assignment Grouping Tests.....	17
3. Control Split Testing	20
F. RESPONDENT RELATIONS	23
1. Background.....	23
2. Interviewer Questionnaire	23
3. Respondent Questionnaire	25
G. CONCLUSIONS	26
H. ACKNOWLEDGEMENTS	28
APPENDIX A: RESPONSE RATE DEFINITIONS.....	29
APPENDIX B: EXPLANATION OF STATISTICAL TESTS	31
APPENDIX C: CONTROL SPLIT DATA: REGIONAL BREAKDOWN	35
REFERENCES	39

ELECTRONIC PUBLICATIONS AVAILABLE AT
www.statcan.ca



A. EXECUTIVE SUMMARY

A test of incentives was done on the 1997 Survey of Household Spending. The purpose of the test was twofold: 1) to gauge what effect incentives would have on respondent relations and 2) to determine if incentives have a positive effect on response rates. This paper deals mainly with the second purpose.

The two incentives tested were: i) a one-year subscription to the quarterly Statistics Canada publication Canadian Social Trends and ii) a telephone calling card good for 20 minutes long distance calling anywhere in North America.

For the experimental design, each interviewer assignment was randomly designated as either magazine or telephone card, ensuring an even split across the regions. Then, within each assignment, all clusters were designated as either incentive or control in a 60%-40% split. This ensured that each interviewer would work with only one incentive, that each cluster would be in the same incentive group, and that the final split of control-magazine-card would be approximately 40%-30%-30%.

Different statistical tests performed at national and regional (Atlantic, Quebec, Ontario, Prairies, and BC) levels reveal the following:

- i) At first blush, it would appear that the magazine had a significant effect at both the national level and in the Atlantic region, as well as in cities of 100,000 to 500,000.
- ii) The magazine assignments differed significantly (with respect to response rates) from the telephone card assignments irrespective of incentive. That is to say there was significant evidence of a difference between the control groups of the two sets of assignments. This difference was most prevalent in the Atlantic region.
- iii) After controlling for assignment effects, i.e. comparing the response rates between an incentive group and the portion of the control group that had the same set of interviewers, it was found that there was no evidence of any effect for the magazine at the national or regional levels. However, the telephone card had a significant effect on response rates in the Atlantic and BC regions as well as at the national level.
- iv) In the Atlantic, the response rate in the telephone card group was 2.2 percentage points higher than the control (83.3% compared to 81.1%); in BC it was 3.5 percentage points higher (76.6% compared to 73.1%), while at the national level the increase was 1.8% (82.3% compared to 80.5%).

There was no significant effect for the telephone card in the other 3 regions.

- v) The above results are consistent with a separate study that analysed the effect of incentives on respondent relations. This study showed that while respondents and interviewers had generally positive reactions to both incentives, the telephone card was the preferred incentive for both respondents and interviewers.

Overall, it is concluded that using telephone cards as an incentive will have a positive effect on respondent relations and will slightly increase response rates. The high cost of giving such an incentive to all responding households must be weighed against these moderately positive effects in any decisions involving incentives.

B. 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. The final response rate was only 77% in 1996, despite being mandatory.

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 would be collapsed in an attempt to shorten the length of the survey and decrease response burden. Finally, it was decided to try incentives once again as a method of improving response rates. This was not the first time incentives were used: a study to test whether giving the respondents incentives would increase the response rate of the survey was done for the 1990 FAMEX with the results being largely inconclusive.

This document will outline the experimental design and the final analysis of the results. These results will be compared with a separate study done jointly by the Income Statistics Division (ISD) and the Survey Operations Division (SOD) on respondent and interviewer reactions to the incentives. It will also look at other analyses done on incentives for other surveys as well as that done on FAMEX in 1990.

C. BACKGROUND

1. Why Incentives?

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.

2. Tests of Incentives for Other Surveys

Before going into the analysis of our test results we will look at some other results that have come out of testing incentives in other surveys.

In the National Consumer Expenditure Survey conducted in 1972-1973, the U.S. Bureau of the Census tested the effect of giving \$5.00 cash as an incentive for participation. At the same time they tested the effect of different collection methods: filling in a diary, receiving daily telephone calls, or the respondents choice of the two methods. The sample size was quite small—only 409 households in and around Chicago divided into city and suburban households. Despite this small sample, significant results were achieved. Although the incentive had no significant effect for suburban households, the incentive group in the city showed a great increase in response rate (85% to 68%) which resulted in a significant difference at the overall level (85% to 73%). They also found some evidence that responding incentive households tended to have better data quality than responding non-incentive households. Their overall conclusions are that the offer of compensation for this survey would have highly favourable effects. For further details of this study see [7].

In 1985, the U.S. Public Health Service did some testing in preparation for their 1987 National Medical Expenditure Survey. One of the focuses of these tests was to compare the effects of prepaid and promised incentives on response rates in a mail-in survey. They found that the response rate of those who had been given a \$5.00 cheque with their questionnaire responded at a significantly higher rate than those who received a note promising a \$5.00 cheque and those who received no promise of payment (the rates in the three groups were 73%, 60%, and 66% respectively). They infer that a prepaid payment gives the respondent a subtle obligation to fill out the questionnaire, while the promised payment can give the impression that if you don't want or need the money you need feel no obligation to fill it out. It should be noted that the sample in this study was very small (about 300 households) and it was biased in the sense that it purposely over-represented the poor, elderly, and non-white population to meet sample size requirements for these interest groups. Nonetheless, their results and subsequent inferences are intriguing, and to be fair to the authors they base their conclusions on other studies as well as their own. For further details see [2].

The U.S. Bureau of the Census' longitudinal Survey of Income and Program Participation underwent a redesign for 1996 and with the redesign came a test of incentives to offset an anticipated drop in response rates. They were interested in determining whether \$10 and \$20 incentives would increase both the current response rate as well as the response rate for subsequent interviews for the longitudinal survey. They concluded that the \$20 incentive had a positive effect on response rates at the overall level as well as for both low and high poverty strata. They also found both the \$10 and \$20 incentives to be effective at reducing follow up non-response both at the overall level and in high poverty

strata. This indicates an initial incentive can prevent households from later dropping out of a longitudinal survey. Their results are based on a total sample size of just over 10,000 initial dwellings. For more details of this experiment see [3].

These are just three illustrations of the many previous experiments that have been done to test incentive effects on response rates and other survey aspects. While they can give us some insights into our own experiment, it is also important to note that many factors influence such a test: the incentive, the survey content, the type of interview, the region, and even the survey year can all impact the results of the experiment. We will now look at a previous experiment on the same survey done for a previous year (1990).

3. Incentive Test Performed for FAMEX 1990

The 1990 FAMEX survey was done not for the entire nation, but only for 15 major cities across the country. The sample size was about 7500 dwellings. The results of the incentive test performed for this survey are given in full in [4]. Here the main points and results are outlined.

Two incentives were used in this design: the Statistics Canada publication “A Portrait of Canada” and a clipboard bearing the Statistics Canada logo. The allocation of the incentives was 40% to the control group and 30% to each of the two incentive groups.

The results were somewhat inconclusive. The authors used Pearson’s chi-square statistics to analyse within city response rates and the sign test to analyse between city effects. Finally they used a logit model and Wald statistics to perform a more detailed analysis, controlling for region, interview, and incentive effect. Their final conclusion is that neither incentive had any effect at the national level, although local effects were observed in some cities.

This earlier study differs from our present study in a number of ways. It used different incentives on an old version of the present survey that was periodically performed only in certain selected cities. The assignment of the incentives was such that every interviewer’s assignment was entirely designated in one group (i.e. an interviewer would give all his or her households the same incentive (or nothing) depending on how the assignment was designated). While this is operationally convenient, it prevented the controlled analysis that was done among interviewer assignments for the 1997 study (see section E2, Assignment Grouping Tests). With the present survey’s coverage being almost all of Canada, an urbanization level variable that will be seen to have an effect on response rates can also be added. The hope was that more conclusive results than those of this 1990 study would be achieved.

D. 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 would be assigned to one of the three incentive groups in a split of 40% to control and 30% to each of the two incentives. Also, for operational reasons an interviewer would work 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 neighbour got a gift for answering your survey and I didn’t”) every dwelling within a cluster would be placed in the same incentive 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 incentive or control making the split as close to 60%-40% 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 with a respondent when the offer of incentive could be made were considered for this study. See appendix A for a more detailed description of which dwellings were included.

E. ANALYSIS

1. Analysis of raw data at various levels

The raw data that came out of the experiment was analysed 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—see appendix B for details of this test. 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.

The p-values given can roughly be interpreted as the probability, assuming no incentive effect and all marginal totals are fixed, that the incentive response rate is as high or higher than it actually is. The lower the probability the more likely it is that the incentive is having a positive effect on the response rate. In the statistical community a p-value of around 0.05 or less is generally considered significant evidence that the alternative hypothesis (in this case, the incentive having a positive effect on response rates) is true.

We start by looking at the data at the national level.

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	

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.

It would be naive at this point to assume that the magazine had a significant effect on response rates. Many confounding factors could account for the difference shown above. Since it is not unreasonable to assume that respondents could react differently to incentives in different regions and in different urbanization levels, response rates were computed for each of these two factors.

Splitting the respondents into the five main regions of Canada gives the following results:

Table 2: Regional Breakdown of Response Rates

<u>Region</u>	<u>Incentive</u>	<u>Response</u>	<u>Non Response</u>	<u>Total</u>	<u>Response Rate</u>	<u>P-Value</u>
ATLANTIC	Control	2315	455	2770	83.6	
	Magazine	1765	283	2048	86.2	0.007
	Phone Card	1674	336	2010	83.3	0.621
	Total	5754	1074	6828	84.3	
QUEBEC	Control	1097	187	1284	85.4	
	Magazine	778	136	914	85.1	0.606
	Phone Card	786	131	917	85.7	0.453
	Total	2661	454	3115	85.4	
ONTARIO	Control	962	296	1258	76.5	
	Magazine	740	205	945	78.3	0.167
	Phone Card	682	192	874	78.0	0.214
	Total	2384	693	3077	77.5	
PRAIRIES	Control	2073	393	2466	84.1	
	Magazine	1637	307	1944	84.2	0.465
	Phone Card	1527	282	1809	84.4	0.396
	Total	5237	982	6219	84.2	
BRITISH COLUMBIA	Control	1017	327	1344	75.7	
	Magazine	710	220	930	76.3	0.375
	Phone Card	728	223	951	76.6	0.331
	Total	2455	770	3225	76.1	

We observe 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.

It is also interesting to note the discrepancy in response rates among the regions. The Atlantic, Quebec, and Prairie regions have response rates in the mid 80s while in Ontario and BC the response rates are almost 10 percentage points lower, closer to the mid 70s.

Another way to partition the sample is by urbanization level. The tables have been divided into three sections: those from cities of greater than 500,000, those from cities between 100,000 and 500,000, and those from cities, towns, and rural areas of less than 100,000.

Table 3: Urbanization Level Breakdown of Response Rates

Urbanization Level	Incentive	Response	Non Response	Total	Response Rate	P-Value
1 (> 500,000)	Control	2408	639	3047	79.0	
	Magazine	1723	445	2168	79.5	0.361
	Phone Card	1812	443	2255	80.4	0.125
	Total	5943	1527	7470	79.6	
2 (between 100,000 and 500,000)	Control	1341	378	1719	78.0	
	Magazine	901	198	1099	82.0	0.006
	Phone Card	1063	267	1330	79.9	0.108
	Total	3305	843	4148	79.7	
3 (< 100,000)	Control	3724	641	4365	85.3	
	Magazine	3014	508	3522	85.6	0.384
	Phone Card	2524	454	2978	84.8	0.757
	Total	9262	1603	10865	85.2	

Here we observe that the response rate for the magazine is greater than that of the control group for all three urbanization levels. The difference is highly significant for the middle level. For the phone card, response rates improved over control in the two higher level urbanization levels, but not significantly so. The rates were lower than control in level 3.

Again note the difference in response rates among the groups. While the levels 1 and 2 have almost identical total response rates, level 3 has a much higher rate (between 5 and 6 percentage points higher).

Since the Atlantic region was the only one that showed significance for either of the incentives, it will be broken down by urbanization level to see if anything can be learned about which urbanization group is responding better to the magazine. Since there are no cities in the Atlantic region of more than 500,000 people, the region can be broken down into two groups as shown below:

Table 4: Urbanization Level Breakdown of Response Rates in Atlantic Region

		<u>Incentive</u>	<u>Response</u>	<u>Non Response</u>	<u>Total</u>	<u>Response Rate</u>	<u>P-Value</u>
2 (between 100,000 and 500,000)	Control		700	208	908	77.1	
	Magazine		425	81	506	84.0	0.001
	Phone Card		630	157	787	80.1	0.078
	Total		1755	446	2201	79.7	
				<u>Non</u>			
3 (< 100,000)	Control		1617	247	1864	86.7	
	Magazine		1342	202	1544	86.9	0.463
	Phone Card		1045	179	1224	85.4	0.872
	Total		4004	628	4632	86.4	

We see that the increased response rate in the magazine group for the Atlantic region can be attributed almost entirely to the cities of over 100,000 people. In the other group the response rates for control and magazine are almost identical (86.7 and 86.9). The telephone card does well in the large cities (just over significance level), while it does poorly in the smaller areas.

The results of the test for magazine effect in the larger Atlantic region cities are highly significant (p-value of 0.1%). This leads us to wonder about the allocation of incentives in those areas. We will return to this point later on in the section.

All these results were somewhat puzzling based on other information that was available. The results of the respondent relations study showed that both respondents and interviewers were having much more favourable reactions to the telephone card than they were to the magazine—see section F. 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, 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 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 interview 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* quantifies this difference and can determine whether it is significant.

Alternatively we could look at the values of the differences of proportion and take their ordinance into account for our test (something the sign test does not do). All the proportion differences for each interviewer assignment are examined and their magnitudes ranked from lowest to highest. The Mann-Whitney test (or Wilcoxon

signed rank test) shows how different these observed ranks are from what might be expected under a null hypothesis of no incentive effect.

Finally the values of the differences of proportion themselves (rather than just their ranks) can be looked at. A paired t-test can be performed, again to see if there is any significant evidence against the null hypothesis of no incentive effect.

It should be noted that the three tests are given in order from most to least robust. The sign test requires no assumptions and is always valid. The Mann-Whitney test assumes that the distribution of the proportions is symmetric about 0—this would be a reasonable assumption under the null hypothesis. The paired t-test assumes the distribution of the proportions is normal—again not an unreasonable assumption under the null hypothesis. Further details about these tests can be found in Appendix B.

The table below 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.

Table 5: Assignment Grouping Tests at Canada 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>
Canada	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

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 the above table and check the results. (Note that interviewer assignments can contain many different urbanization levels; thus analysis at that level could not be performed.)

Table 6: Assignment Grouping Tests at Regional Level

					<u>P-Values</u>		
	<u>Incentive</u>	<u>Positive Difference</u>	<u>Negative Difference</u>	<u>Total</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
QUEBEC	<u>Incentive</u>	<u>Positive Difference</u>	<u>Negative Difference</u>	<u>Total</u>	<u>Sign Test</u>	<u>Mann-Whitney</u>	<u>Paired-t</u>
	Magazine	26	30	56	0.748	0.724	0.668
	Phone Card	26	29	55	0.705	0.579	0.487
Total	52	59	111	0.776	0.712	0.614	
ONTARIO	<u>Incentive</u>	<u>Positive Difference</u>	<u>Negative Difference</u>	<u>Total</u>	<u>Sign Test</u>	<u>Mann-Whitney</u>	<u>Paired-t</u>
	Magazine	31	30	61	0.500	0.318	0.348
	Phone Card	36	25	61	0.100	0.344	0.451
Total	67	55	122	0.160	0.269	0.359	
PRAIRIES	<u>Incentive</u>	<u>Positive Difference</u>	<u>Negative Difference</u>	<u>Total</u>	<u>Sign Test</u>	<u>Mann-Whitney</u>	<u>Paired-t</u>
	Magazine	47	39	86	0.225	0.493	0.495
	Phone Card	40	42	82	0.630	0.760	0.763
Total	87	81	168	0.350	0.682	0.699	
BRITISH COLUMBIA	<u>Incentive</u>	<u>Positive Difference</u>	<u>Negative Difference</u>	<u>Total</u>	<u>Sign Test</u>	<u>Mann-Whitney</u>	<u>Paired-t</u>
	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 quite 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, fails 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 for. 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 would elicit the same response rate as those assigned to the telephone card group is not valid these tests will still be valid. In the next section this assumption will be 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 7: Control Split Tests at Canada Level

Magazine

<u>Incentive</u>	<u>Response</u>	<u>Non 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	

Phone Card

<u>Incentive</u>	<u>Response</u>	<u>Non 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

Comparing totals: p-value = **0.004**

The results are quite different when the control group is split. Here 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 below the table.

The three p-values shown here 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. The last p-value, comparing the entire samples (magazine interviewer’s response rate of 82.9% with telephone card interviewer’s response rate of 81.6%), is also significant at 0.004, but the previous arguments clearly show that this difference is not due to the incentive but to some interviewer effect.

We next broke down the above table 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 below—the complete table can be found in Appendix C.

Table 8: 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
	Telephone Card	Incentive	86.2		
Quebec	Magazine	Control	81.1	0.057	0.267
	Telephone Card	Incentive	83.3		
Quebec	Magazine	Control	86.2	0.742	0.267
	Telephone Card	Incentive	85.1		
Ontario	Magazine	Control	84.8	0.330	0.554
	Telephone Card	Incentive	85.7		
Ontario	Magazine	Control	76.4	0.204	0.554
	Telephone Card	Incentive	78.4		
Prairie	Magazine	Control	76.6	0.277	0.808
	Telephone Card	Incentive	78.0		
Prairie	Magazine	Control	82.9	0.184	0.808
	Telephone Card	Incentive	84.2		
BC	Magazine	Control	84.4	0.458	0.013
	Telephone Card	Incentive	84.4		
BC	Magazine	Control	78.5	0.852	0.013
	Telephone Card	Incentive	76.4		
BC	Magazine	Control	73.1	0.060	0.013
	Telephone Card	Incentive	76.6		

We see that 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. 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% 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.

F. RESPONDENT RELATIONS

In addition to looking at the effect the incentives had on survey response rates, a simultaneous study was done to evaluate incentive impact on respondent relations. The results of this study are outlined in full in [6] but are summarized here and compared to the response rate results.

1. Background

To gauge respondent reaction, all households that received one of the two incentives were given a short questionnaire to fill out that asked questions related to their thoughts about the incentive. This questionnaire was left with the respondent once the interview was complete and the respondent was asked to fill it out and mail it to head office. In order to reduce the already heavy burden placed on all SHS respondents, no follow up was done for the non-response to this incentive questionnaire.

Each interviewer was also asked to fill out a questionnaire asking about their opinion on the incentive's effect, respondent reaction to it, and how comfortable they were in offering it.

Some of the results of these two questionnaires will be compared with the response rate results obtained in the previous section.

2. Interviewer Questionnaire

After completing all their SHS interviews, each interviewer was asked to fill out a short questionnaire regarding their opinions and level of comfort in offering incentives to the respondents. The questionnaire was to be mailed to head office

upon completion. A total of 683 out of 785 questionnaires were received for a response rate of 87%.

Recall that each interviewer worked with only one incentive and it has been determined that there was likely some bias in the assignment of the incentives to interviewers. This should be considered when analysing the data below.

Table 9: Data from Interviewer questionnaires

%	Canada	Atlantic	Quebec	Ontario	Prairies	BC
Comfortable with Incentive						
Magazine	72	77	81	65	65	72
Telephone Card	89	85	96	97	84	82
Yes, I would like to give this incentive again in the future						
Magazine	44	49	60	33	33	47
Telephone Card	83	86	94	86	76	81

Incentive was generally appreciated	
Magazine	49
Telephone Card	63
Incentive had Positive Effect on Response Rates	
Magazine	11
Telephone Card	29

Table 9 shows that the interviewers were generally comfortable with offering the incentive, with the telephone card having a higher comfort level in every region. This difference is even greater when we look at the percentage of interviewers who would like to offer the incentive in the future. 83% wanted to offer the telephone card again, while only 44% wanted to offer the magazine. Again the telephone card superiority is prevalent in every region.

The second part of the table gives an indication of the interviewers impressions of respondents' reactions to the incentive. The first column tells us that 49% of magazine interviewers felt that their incentive was generally appreciated by respondents compared to a 63% rate for the telephone card. Similarly, only 11% of magazine interviewers felt it had a positive effect on the response rate, compared to 29% of the telephone card interviewers.

Clearly from the point of view of the interviewers, the telephone card is the preferred incentive. We will now look at the respondent feedback questionnaire titled *We Were Wondering*.

3. Respondent Questionnaire

Before we compare the response rate results to the respondent questionnaire results, it is important to note the difference between the two samples. When we look at response rates, our sample consists of every dwelling in the SHS sample where contact with a respondent was made. For the respondent questionnaire our target population is only those respondents in the incentive group that responded to the SHS questionnaire, and our sample is only those who filled out and returned the incentive questionnaire. Thus rather than an impression of how the population in general views the incentive, we get an impression of how those who do actually respond view the incentive. There is likely much bias in the questionnaire as well since it is not unreasonable to assume that those who bothered to fill out the incentive questionnaire are not the same with regards to opinions on the incentive as those who didn't bother. These considerations should be taken into account in all the comparisons done below.

The *We were wondering* incentive questionnaires had a response rate of about 38% nationally and all results are based on these 4160 questionnaires that Statistics Canada received back from SHS respondents. The table below summarizes some of the results.

Table 10: Data from *We were wondering* questionnaires

%	Canada	Atlantic	Quebec	Ontario	Prairies	BC
Appreciated Incentive						
Magazine	79	82	77	77	75	80
Telephone Card	95	96	93	95	95	97
Influenced Participation						
Magazine	19	21	21	16	17	10
Telephone Card	18	21	14	17	20	11

Table 10 shows the percentage of respondents on the incentive questionnaire that indicated they appreciated the incentive as well as the percentage that indicated the incentive influenced their decision to participate in the survey. The table is broken down by region and by incentive.

There is quite a difference between the rates of appreciation for the two incentives that is consistent across all regions. Fully 95% of the responding telephone card people said they appreciated the gift while only 79% of magazine respondents appreciated theirs. The higher appreciation rate for the telephone card is not unexpected since almost everyone can find use for a long distance telephone card, while all people do not necessarily appreciate a subscription to a Statistics Canada publication. The high numbers overall are a good indication that both incentives were well received by respondents.

The table also shows that just under 20% of incentive respondents felt the incentive influenced their decision to participate. This result is rather curious for several reasons. First, while the telephone card was generally appreciated more, the two incentives were about equal in the participation influence category. But more interestingly, one of the final questions on the questionnaire asked respondents their main reasons for participating in the SHS. Several choices were given and respondents could fill in more than one answer. Only 3.4% (141 out of 4087) filled in the received incentive as one of the *main reasons* for participating while even less, only 0.5% (11 out of 2154), had it as their sole answer—a far cry from the 18% who said the incentive *influenced* their participation. Obviously many respondents interpreted the questions differently—likely many thought the term “main reason” in the final question meant “was it the difference between your responding or not responding”, while not having this same interpretation of “influenced”. The response rate data looked at in the previous section obviously supports this reasoning. If 18% of respondents in the incentive group truly felt the incentive was the difference between their responding and not responding it would be expected to see an increase of around 14% (18% multiplied by an approximate 80% response rate) from the control group to the incentive groups. Obviously we did not come close to this in any region or breakdown of data. The observed increases are (depending on the incentive and on the region) anywhere from 0% to about 4% which is more in line with the 3.4% reported in the final question (3.4% multiplied by an 80% response rate is about 2.7%).

G. 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 to us 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 these 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 the effect in these two regions.

It is worth noting at this point that the 1997 SHS was conducted with an artificially inflated sample size in the Atlantic region. Since we have concluded that the telephone card effect at the national level is largely attributable to this region, this overrepresentation could conceivably be one of the causes. Also, if the regional response rates are examined, we notice that the Atlantic, Quebec, and Prairie regions all have much higher response rates than the Ontario and BC regions. Since it is strongly desirable to increase the response rates in these two regions, an incentive in BC would be highly effective.

The data collected from the respondents and the interviewers clearly indicates that incentives were generally well received by both respondents and interviewers. It also indicates the telephone card was the more appreciated of the two. This lends positive reinforcement to our response rate results and shows that a potentially better relationship with respondents and better image for Statistics Canada in general are further benefits that can be gained by using incentives in the SHS.

There remains room for further study. We have shown that there was some non-incentive related difference between the two sets of interviewer assignments, but what is the underlying cause of this difference? The two possibilities seem to be either or both of 1) the magazine interviewers were, on average, superior to the

telephone card interviewers or 2) The areas assigned to the magazine interviewers were, on average, more co-operative and higher responding areas. The fact that our sample size was quite large and the two incentives were assigned to the areas completely at random (whereas the interviewers were assigned by the regional offices after the incentive areas were designated) makes 1) seem more probable than 2). Unfortunately, complete interviewer profiles (i.e. gender, age, years of experience etc.) were not available for such an analysis at the time of writing. If this information does become available, perhaps the reason for the difference could be determined. The size of the assignment is a possible third factor that could also have had an effect on this difference (i.e. a large assignment could be burdensome to an interviewer and cause decreased response rate).

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.

H. 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.

APPENDIX A: RESPONSE RATE DEFINITIONS

After visiting each household in the original sample of the 1997 SHS, the interviewer would give the household one of the following eligibility codes:

00	useable and in balance
04	useable and out of balance
07	Incomplete expenditures
08	Incomplete income
09	Both incomplete income and expenditures
17	Members ineligible
22	Dwelling demolished
23	Dwelling under construction
24	Dwelling vacant
30	Household refusal
34	No contact
39	Temporary absence
49	Refusal, at least Sections A & B complete
52	Unusual circumstances

Those who were assigned the code 00 or 04 were considered for our purposes to be *responses*.

Those who were assigned the code 07, 08, 09, 30, or 49 were considered for our purposes to be *non-responses* (codes 07, 08, and 09 implied the respondent refused to answer a significant portion of the questionnaire).

Those who were assigned the code 34, 39, or 52 were considered for our purposes to be *no-contacts*. Since these respondents were never given the opportunity to refuse or participate in the survey, they do not reflect the effect the incentives had on response rates and are not counted in any of the rates that are given. The codes 17 (ineligible) and 22, 23, and 24 (vacant) are also not included in any calculations. Thus for this analysis *response rate* is defined as:

$$\frac{\text{responses}}{\text{responses} + \text{nonresponses}}$$

Where responses and non-responses are as just previously defined. This quantity will normally be multiplied by 100 and expressed as a percent.

This concept of response rate differs from the normal Statistics Canada concept. Usually the response rate is reported as the number of response divided by the number of eligible dwellings in the sample. That is to say the no-contact codes would normally be included as non-response in the calculation. They were excluded here for the reasons mentioned above. Thus the response rates reported in this document that are of interest for our incentive study are higher than those reported in official documents covering the 1997 SHS.

APPENDIX B: EXPLANATION OF STATISTICAL TESTS

Fisher's Exact Test

If one makes k draws from a jar containing m red balls and n black balls then the probability that one draws exactly x red balls is:

$$P(X = x) = \frac{\binom{m}{x} \binom{n}{k-x}}{\binom{m+n}{m}}$$

This is the well-known hyper-geometric distribution. This distribution can be used to test one or two-sided hypotheses for two by two contingency tables. An example from the paper will be used to illustrate this. Since only one-sided hypotheses were used in this paper this will be demonstrated.

<u>Incentive</u>	<u>Response</u>	<u>Non</u> <u>Response</u>	<u>Total</u>
Control	3583	866	4449
Phone Card	5399	1164	6563
Total	8982	2030	11012

This is the control group split data for the telephone card at the national level. We wish to test the null hypothesis H_0 : the telephone card has no effect on response rates versus the alternative hypothesis H_a : the telephone card increases the response rate.

Let us assume all the marginal totals are fixed. That is to say of the 11,012 dwellings, we know 4449 were in the control group and 6563 were in the incentive group. We also know that 8982 responded and 2030 were non-responses. It can be shown that the expected number of incentive responses under the null hypothesis is $(8982 \times 6563)/11012 = 5353$. If the alternative hypothesis is true, we would expect to have more responses in the incentive group and obtain a number higher than this (as we do in this example). The p-value of the test is the probability under H_0 that the number of incentive responses received is greater than or equal to the observed 5399. This probability can be calculated using the hyper-geometric distribution, if we think of the number of black and red balls in our jar as the number of control and incentive households in the sample (in this case 4449 and 6563) respectively. We can imagine selecting 8982 balls (i.e. responses) and calculating the probability of picking 5399 or more red balls (i.e.

incentive households). Calculating p-values of contingency tables using the hyper-geometric distribution is known as Fisher's Exact Test. The formula for this probability in our example is:

$$P(X \geq 5399) = \sum_{x=5399}^{6563} \frac{\binom{6563}{x} \binom{4449}{8982-x}}{\binom{11012}{6563}}$$

As reported in the original table this probability turned out to be 0.012. For further details on Fisher's Exact Test see [1]. All p-value calculations from Fisher's exact test in this paper were computed using the PROC FREQ command in SAS.

Assignment Grouping Tests

Three different tests were performed on the data after grouping it by interviewer assignment. These tests will be outlined here. We will also more clearly explain what was done to the data to prepare it for these tests.

An interviewer assignment generally contained around 15 to 55 households (after no-contacts, vacancies etc. were removed). These households could be split into those who responded and those who didn't as well as into those who received an incentive and those who didn't. The incentive and control response rates would then be computed. Their difference would be the quantity of interest used in the tests.

To give an example, assignment number 23510 contained 52 dwellings of which 39 were of interest to us (i.e. eligible dwellings that the interviewer made contact with). Of the 21 households in the incentive group, 18 responded and 3 did not for an incentive response rate of $18/21 = 0.857$. In the control group there were 18 households of which 13 responded and 5 did not for a control response rate of $13/18 = 0.722$. Thus the difference in the response rates is $0.857 - 0.722 = 0.135$.

After getting this difference for all 700+ interviewer assignments, the following three tests can be performed.

Sign Test

In performing the sign test, we ignore the magnitude of the differences and just look at the sign, i.e. was the difference positive or negative. Under a null hypothesis of no incentive effect we would expect there to be about an equal number of positive and negative differences. A p-value of the observed number of positive differences x can be calculated using the binomial distribution:

$$P(X \geq x) = \sum_{k=x}^n \binom{n}{k} \left(\frac{1}{2}\right)^n$$

Where n is the number of interviewer assignments in the sample. For example, for the telephone card assignments at the national level we observed 188 positive differences and 163 negative differences. Thus for the formula above, $x=188$ and $n=188+163=351$. The p-value in this case turned out to be 0.100. For more details on the sign test see [1] or [5]. All p-values for the sign test were calculated using the QBINOM command in S-PLUS.

Mann-Whitney Test

Instead of just looking at the signs of the differences, we can take into account their ordinality as well. We look at the magnitudes of each difference (ignoring sign) and order them from smallest to largest. We then rank them from 1 (the smallest) to n (the largest) where n is the number of assignments we have. We now separately sum the ranks of the positive and negative differences and compare them. Under the null hypothesis of no incentive effect we would expect these sums to be about the same. Under the alternative hypothesis of a positive incentive effect we would expect the sum of the ranks of the positive differences to be higher than the same sums of the negative differences. The p-value of the test reveals how large this difference is using the Wilcoxon test statistic. The formulas involved here are rather complex and will not be shown, but are fully detailed in [5]. All Mann-Whitney p-values were calculated using the WILCOX.TEST function in S-PLUS.

Paired t-tests

In this test we use the actual values of the differences to compute a test statistic. Under very general assumptions, it can be shown that under the null hypothesis these differences will follow the t-distribution with a mean of 0. We calculate the mean and variance of our sample of differences and compare with a Student's t-statistic. The p-value here represents how much greater than 0 our mean actually is. For full details see [5]. All t-test p-values were calculated using the T.TEST function on S-PLUS.

**APPENDIX C: TABLE OF REGIONAL BREAKDOWN OF CONTROL
SPLIT DATA**

Table 11: Control Split Tests at Regional Level

Atlantic

Magazine

<u>Incentive</u>	<u>Response</u>	<u>Non Response</u>	<u>Total</u>	<u>Response Rate</u>	<u>P-Value</u>
Control	1245	205	1450	85.9	
Magazine	1767	283	2050	86.2	0.408
Total	3012	488	3500	86.1	

Phone Card

<u>Incentive</u>	<u>Response</u>	<u>Non Response</u>	<u>Total</u>	<u>Response Rate</u>	<u>P-Value</u>
Control	1072	250	1322	81.1	
Phone Card	1675	336	2011	83.3	0.057
Total	2747	586	3333	82.4	

Comparisons between magazine and phone card assignments

2 Control Groups: p-value = **0.000**

2 Incentive Groups: p-value = 0.006

Comparing totals: p-value = **0.000**

Quebec

Magazine

<u>Incentive</u>	<u>Response</u>	<u>Non Response</u>	<u>Total</u>	<u>Response Rate</u>	<u>P-Value</u>
Control	530	85	615	86.2	
Magazine	778	136	914	85.1	0.742
Total	1308	221	1529	85.5	

Phone Card

<u>Incentive</u>	<u>Response</u>	<u>Non Response</u>	<u>Total</u>	<u>Response Rate</u>	<u>P-Value</u>
Control	569	102	671	84.8	
Phone Card	786	131	917	85.7	0.330
Total	1355	233	1588	85.3	

Comparisons between magazine and phone card assignments

2 Control Groups: p-value = 0.267

2 Incentive Groups: p-value = 0.665

Comparing totals: p-value = 0.451

Ontario

Magazine

<u>Incentive</u>	<u>Response</u>	<u>Non Response</u>	<u>Total</u>	<u>Response Rate</u>	<u>P-Value</u>
Control	477	147	624	76.4	
Magazine	742	205	947	78.4	0.204
Total	1219	352	1571	77.6	

Phone Card

<u>Incentive</u>	<u>Response</u>	<u>Non Response</u>	<u>Total</u>	<u>Response Rate</u>	<u>P-Value</u>
Control	488	149	637	76.6	
Phone Card	682	192	874	78.0	0.277
Total	1170	341	1511	77.4	

Comparisons between magazine and phone card assignments

2 Control Groups: p-value = 0.554

2 Incentive Groups: p-value = 0.457

Comparing totals: p-value = 0.474

Prairies

Magazine

<u>Incentive</u>	<u>Response</u>	<u>Non Response</u>	<u>Total</u>	<u>Response Rate</u>	<u>P-Value</u>
Control	982	202	1184	82.9	
Magazine	1640	307	1947	84.2	0.184
Total	2622	509	3131	83.7	

Phone Card

<u>Incentive</u>	<u>Response</u>	<u>Non Response</u>	<u>Total</u>	<u>Response Rate</u>	<u>P-Value</u>
Control	943	177	1120	84.2	
Phone Card	1527	282	1809	84.4	0.458
Total	2470	459	2929	84.3	

Comparisons between magazine and phone card assignments

2 Control Groups: p-value = 0.808

2 Incentive Groups: p-value = 0.577

Comparing totals: p-value = 0.744

British Columbia

Magazine

<u>Incentive</u>	<u>Response</u>	<u>Non Response</u>	<u>Total</u>	<u>Response Rate</u>	<u>P-Value</u>
Control	500	137	637	78.5	
Magazine	711	220	931	76.4	0.852
Total	1211	357	1568	77.2	

Phone Card

<u>Incentive</u>	<u>Response</u>	<u>Non Response</u>	<u>Total</u>	<u>Response Rate</u>	<u>P-Value</u>
Control	511	188	699	73.1	
Phone Card	729	223	952	76.6	0.060
Total	1240	411	1651	75.1	

Comparisons between magazine and phone card assignments

2 Control Groups: p-value = **0.013**

2 Incentive Groups: p-value = 0.564

Comparing totals: p-value = 0.085

REFERENCES

1. Agresti, Alan (1990): *Categorical Data Analysis*. New York: Wiley.
2. Berk, Marc L., Mathiowetz, Nancy A., Ward, Edward P., White, Andrew A. (1987): *The Effect of Prepaid and Promised Incentives: Results of a Controlled Experiment*. *Journal of Official Statistics*, No. 4, pp. 449-457.
3. James, Tracy L. (1997): *Results of Wave 1 Incentive Experiment in the 1996 Survey of Income and Program Participation*. *Proceedings of the Survey Research Section of the American Statistical Association*.
4. 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.
5. Lehmann, E.L. (1975): *Nonparametrics: Statistical Methods Based on Ranks*. San Francisco: Holden-Day, Inc.
6. 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.
7. Sudman, Seymour and Ferber, Robert (1974): *A Comparison of Alternative Procedures for Collecting Consumer Expenditure Data for Frequently Purchased Products*. *Journal of Marketing Research*, Vol. XI pp 128-135.