On the Validity of Online Panels

Chuck Chakrapani
Ted Rogers School of Management, Ryerson University

In recent years there have been heated discussions about the validity of online panels as compared to telephone interviews. Several empirical studies have been carried out to assess the superiority of one method over the other. This paper discusses three factors that potentially affect the validity of methods: response rate, coverage and the need for external validity. It suggests that the effects of coverage and response rates can be very similar on the final results. Because the nature of both telephone surveys and online panels are dynamic and constantly changing, we need to consider when to do what and not be overly concerned about establishing the superiority of one method over another.

Introduction

Whenever a new methodology comes into existence, it creates excitement because of its novelty and potential. At the same time, it is also subject to severe skepticism and scorn because of its potential weaknesses. Online research is no different. After considerable discussion (for example, Deutskens *et al.* 2004; Eggars 2006; Halpenny 2007; Halpenny & Ambrose 2007; Miller 2007; Mills 2007), some of which has been very heated, it may be time to view the debate dispassionately in the light of the best evidence available to us. The fact that online research currently has some egregious weaknesses such as coverage error in some segments can be supported with statistics. Yet online research is a rapidly changing landscape. We need to explore not only the implications of coverage error but also whether online research can be used in spite of its perceived weaknesses, rather than dismissing it summarily because of them.

A good place to start is to compare the online method with a method it is replacing: telephone interviewing. It is interesting to note that when CATI systems started replacing personal interviews some 20 or 30 years ago, many were skeptical about telephone interviewing. Wouldn't it be easier for the respondent to refuse? Wouldn't it be easier for respondents to give untruthful answers because "it is just a voice on the other side of the phone"? Wouldn't it be easier for the respondent not to pick up the phone at all? Wouldn't potential respondents screen all calls using caller ID function? Those questions have not gone away completely and yet now, telephone interviews are well accepted and questions are being raised about online interviewing.

In this paper, I deal with the general issues raised by online method vs. telephone method, and not necessarily with the nuances of sample selection procedures. I am assuming that telephone interviews use some type of probability sampling and online interviews use a properly recruited and maintained panel or a reliable customer database. This paper intentionally confines itself to a very narrow issue: given the problems with coverage and related issues, should we continue to use online panels or should we simply revert back to doing mainly telephone surveys?

Three Core Issues

There are many issues that influence the choice of telephone over online surveys or vice versa. Of these, three can be considered core issues: coverage error, response rates and the relevance of external validity.

1. Coverage error

The issue: *According to Statistics Canada, 98% of Canadian homes have access to home telephone. But only 68% of Canadian homes have access to the Internet. So we systematically exclude about one-third of the population. This can distort results when we collect information using online surveys.*

One of the fundamental assumptions of sampling is that everyone in the population (the target audience) should have an equal (or at least a measurable) opportunity of being included. This ideal is seldom, if ever, achieved in marketing research. In Canada, for example, some areas such as NWT have been typically excluded when choosing a sample. Respondents who may be eligible are excluded because they are institutionalized for various reasons. The objective here is not to achieve full coverage, but to have reasonable confidence that the achieved coverage is good enough and not likely to distort the results. Our assumption here is that the part of the population we intentionally excluded is similar to the part

of the population from which we drew the sample, and therefore our results are generalizable to the entire population.

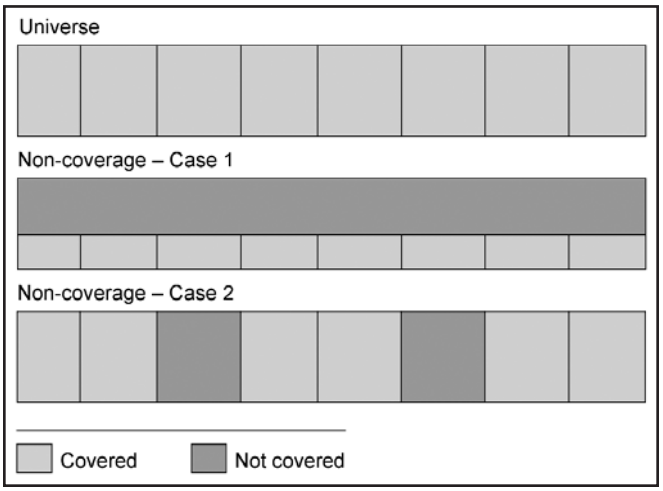
The most prominent example of coverage error in marketing research related to the poll conducted by *Literary Digest* some 70 years ago (Chakrapani & Deal 2005). The *Literary Digest* poll of voters had a sample size of 2.4 million, an astonishing (by today's standards for mail surveys of the general population) 24% return on the 10 million surveys it sent out. The voter population of the entire US was only 78 million then. Given that 2.4 million of the potential 78 million participated in the survey, one would expect the poll to be remarkably accurate. The poll predicted that presidential candidate Landon would receive 57% of the votes and win in a landslide. In the election, Landon received only 38% the votes and Roosevelt won in a landslide. A sample of 2.4 million would have almost no margin of error, and yet the poll was wrong by nearly 20 percentage points. The problem was coverage error. *Literary Digest* got its respondent lists from automobile and telephone owners (less universal about 70 years ago), country club memberships, its own subscribers, etc. All these lists excluded less affluent Americans who tended to vote Democrat (Roosevelt's party), thereby severely skewing the polls in favour of Landon.

On the face of it, it would seem that online interviews are a severely flawed technique compared to telephone interviews. However, all coverage errors do not necessarily have serious implications in practice. If we can assume that those who are covered are similar to those who are not with regard to study variables (the variables that a researcher is interested in), the results may not be subject to much distortion. For example, if we do a survey about detergents over a week and about 10% of the households happen to be on vacation during that period, our coverage may drop without its necessarily introducing any appreciable bias in our findings. We have no reason to expect that those who went on holidays during the interview period will be systematically different from those who did not. As we saw above, this is not what happened in the Literary Digest poll: those who were different were so in terms their political attitudes. Those who were covered by the poll did not constitute a random subset of the population, but a biased carving of it.

Not all coverage errors are equal. Conceptually, we can identify two types of coverage errors: non-coverage that is across different segments[1] of the population, and non-coverage that is confined to specific segments of the population (See Exhibit 1). The implications of type of non-coverage could be very different.

In the Case-1 type of non-coverage, although some part of the population is not covered, such non-coverage cuts across all segments of the population. For instance, if one were to interview adult females in a given period, it is likely some part of this population would not be available to be interviewed—perhaps they are on their annual vacation. However, this is likely to happen to a greater or lesser extent in every segment of the adult female population. Unless our survey relates to vacation habits, this type of non-coverage is unlikely to affect our estimates based on the survey.

Exhibit 1: Forms of coverage error



The Case-2 type of non-coverage is potentially more serious. Here entire segments of the population are not covered and therefore the results are subject to confounding errors. This is the type of confounding that occurred in the *Literary Digest* poll—entire segments that would have voted for Roosevelt were not polled.

If we apply this logic to the non-coverage of the Internet, it should be fairly obvious that online coverage is unlikely to be unbiased. It is generally known that online usage is consistently higher among certain groups: those who are younger, those who are relatively more affluent, and those who are comfortable with technology. Thus, for example, an online study that measures how comfortable people are with technology will likely overestimate the comfort level. As a general rule, to the extent our survey variables are related to the coverage factors, the survey may give biased results.

Can the coverage error be eliminated or mitigated by stratified random sampling or by weighting the results? For instance, suppose that in our panel younger people are over-represented. We can assign a lower weight to them so they are not over-represented. Another alternative would be to choose the sample such that each subgroup is in the same proportion as in the population. Unfortunately, this will not necessarily correct the problem. This is shown in Exhibit 2.

[1] Here, the word 'segment' refers to any subgroup of the target audience. When a subgroup is not covered, there is a possibility that the segment not covered may share some common traits that may be of interest to the researcher, as will be illustrated later.

Exhibit 2: Effect of weighting of a poorly covered population (example)

	Population %	Online panel %	"Facebook" Usage %	Unweighted Users %	Weighted Users %
18-24	12	27	60	16.2	7.2
25-44	23	23	40	9.2	9.2
44-64	25	20	30	6.5	7.5
65+	40	30	5	1.5	2.0
Total users (unweighted)				33.4	
Total users (weighted)					25.9

When the data were unweighted, Facebook usage was 33.4% of those we interviewed. By weighting, we arrived at a lower estimate of 25.9%. The effect would be similar if we used proportional sampling instead of weighting. Unfortunately, neither weighting nor PPS (Probability Proportional to Size) sampling can solve the problem of coverage. Both these methods would assume that the calculated results would also apply to the non-covered population which, as this example shows, may not be a reasonable assumption. Only those who have access to the Internet can have potential access to Facebook, and this is 100% of those in the panel. Those who have access to the Internet can have no potential access to Facebook, and this can be far less than 100%. Therefore, we cannot always weight our way out of the coverage problem.

There is one other problem with weighting samples that may not be representative because of poor coverage. We can only weight for variables we believe to be influential. Such variables tend to be demographic, such as age, gender and region. But it is quite conceivable that even if our samples are weighted to match demographic characteristics, they may be unbalanced in less tangible variables such as propensity to buy. The weighted sample may resemble the universe in terms of demographics, but carries no guarantees that it will be representative in other aspects as well. When we miss whole segments (as in Exhibit 1, Case 2), the assumption that missing segments will be similar to non-missing segments will not necessarily hold.

Coverage is also becoming a problem with telephone surveys because of cell-only households, which have been rapidly increasing in recent years, and the problem is likely to get a lot worse (Ambrose, Gray & Halpenny 2008). Exhibit 3 (derived from Ambrose & Gray 2007, Ambrose, Gray & Halpenny 2008) shows that cell-only households have grown three-fold in fewer than three years, both in the US and in Canada. While the proportion is still small, it is growing exponentially, or at least it would seem so judging from current trends.

Exhibit 3: Growth of cell-phone-only households

	Penetration Level (%)	
Year	Canada	US
2003 May	1.9	
2003 Jan. – June		3.2
2004 May	2.5	
2004 Jan. – June		5.0
2004 Dec.	2.7	
2005 Jan. – June		7.3
2005 Dec.	4.8	10.5
2006 Dec.	4.9	14.0

More importantly, cell-phone-only households are not uniformly distributed across all segments. For instance, there are 28% of households that are single-person households. Yet they account for 59% of all cell-only households.

Again, if coverage is such a serious problem with online surveys, we should not expect to find a strong correlation between online survey estimates and phone survey estimates. There are many such analyses which show that the estimates derived from telephone studies are strongly correlated with estimates derived from online panels. (It should be borne in mind that this finding does not cover every single subject matter, a topic we will consider later in this paper.) In the early years of online interviewing, Harris Polls (now Harris Interactive) carried out parallel interviewing on the telephone and on the web. Their experience basically confirmed that while there were areas where the two methods gave different results, for other subject areas, online results were comparable to those obtained in telephone surveys. This was several years ago, but such findings have been replicated in recent years. For example, regression analyses were carried out on online vs. telephone survey concept scores on 80+ General Mills. The correlation between online studies and telephone surveys was very high: 0.96 (Miller 2007).

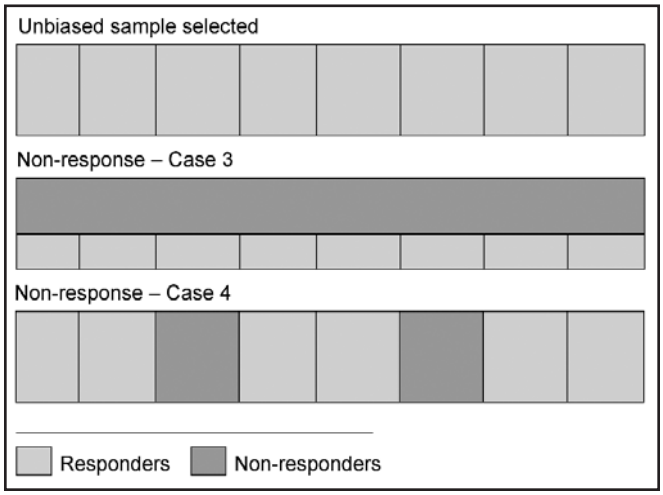
Therefore, even if the coverage advantage currently goes to telephone interviewing, the strong correlation between telephone and online surveys cannot be overlooked. We also need to review another survey research problem: response rate.

2. Response rate

The issue: *Response rate has been declining over the years. According to a MRIA report (2007), only 12% of a sample responds. If 88% of the sample refuses to respond, how can we be certain that the sample has retained its randomness?*

This problem looks unrelated to the first problem of coverage. However, the effect of non-response may not be that different from the effect of non-coverage. In understanding coverage, we asked the question "Is the universe adequately covered by the sample?" In understanding response rate, we ask the question "Is the obtained sample adequately covered by the intended sample?" As in coverage, response rate may come about in two ways, as shown in Exhibit 4.

Exhibit 4: Response rate error



As with coverage, non-response that leaves out entire sub-segments of the intended sample is likely to have more potential repercussion than when the non-response is across all sub-segments. As an example, if we conduct a survey to assess the use of detergents, the response rate may matter less (assuming that a person's refusal to participate in a survey is not related to his or her detergent use). However, if the survey is about estimating the net worth of Canadians and if those who do not respond tend to be those with high net worth, then we would consistently underestimate the net worth of Canadians, an error similar to the coverage error, and similar to the one illustrated by the *Literary Digest* fiasco.

Although the sources of errors are different, both coverage and non-response rate potentially result in an identical problem. *The final sample does not reflect the universe that it is supposed to represent.* The slippage occurs at the universe level for coverage and at the sample level for non-response. We cannot assume without evidence that the slippage at the universe level is more or less serious than slippage at the sample level.

To assess the severity of non-response, researchers often compare results obtained when the response rate is low (after a callback or two) with those obtained when the response rate is higher (after several callbacks). The basic assumption is that if the two sets of results are comparable, then non-response is not a serious issue. For example, Keeter *et al.* (2007) and Halpenny (2007) report that a comparison of specially undertaken studies of similar sample size but different response rates shows no meaningful differences. Exhibit 5 shows one of the findings (a study designed on behalf of MRIA by the Response Rate Committee). The exhibit shows that the results are remarkably stable, whether the response rate is 9% or 31%.

Exhibit 5: Response rates vs. household equipment ownership

	9% RR	31% RR	Sig. Diff
	%	%	
Microwave Oven	95	95	N
Automatic Dishwasher	63	61	N
Gas BBQ	59	57	N
Security System	34	37	N
Espresso/Cappuccino Maker	14	13	N

Such studies are badly needed. They add to our understanding of the effect of response rates on survey results. Unfortunately, they fail to consider one very influential factor: the correlation between study variables and possible differences in estimates between responders and non-responders. Response rate studies cannot be generalized across the board. They can only be generalized for specific products or services. As examples, consider the following scenarios:

1. A researcher wants to know how satisfied customers of a certain brand of refrigerator are and calls a sample of all those who own that brand. The response rate is 15%.

2. Another researcher wants to know the average net worth of all those whose annual income is at least $200,000. The researcher starts with a reasonably complete list of those whose income is in excess of $200,000. The response rate is 15%.

In the first example, we may have no reason to suppose that those who we failed to reach have any systematic reason to evaluate the refrigerator differently than those we did reach. In the second example, we may have reason to believe that the higher the person's net worth, the less likely the interviewer would be to reach that person. Furthermore, the respondent would probably be less likely to agree to answer the survey once he or she is reached. Therefore, in the first example the obtained results may have little bias, while in the second example the bias might lead to substantial underestimating of the net worth of our target audience, even though the response rates in both cases are low and are identical.

Researchers who have considered the effect of drastically lowered response rates (Halpenny 2007; Keeter *et al.* 2007) point out that if the sample is chosen well, the response rate does not really matter. However, this does not conclusively prove that response rate does not matter. Coverage error mattered in *Literary Digest* polling. But does it matter that much when

we use white pages instead of random digit dialing? Probably not, because the bulk of marketing research surveys is concerned with mass market products and services, and a proportion of non-coverage across different segments is likely to distort the data less than non-coverage of entire sub-segments. The same logic holds for non-response rates. Several years ago, as support for CARF standards, I worked out the potential effect of non-response when there is a response rate variation between the designed sample and the obtained sample. (See Exhibit 6.)

mostly come from segments that are critical to survey results, then a bias occurs. This is what happened in the *Literary Digest* poll. One could also speculate whether some of the results reported by Halpenny (2007) could be attributed to this effect (see Exhibit 7). Could it be that non-responders are more likely to have a credit card than responders? Could it be that harder-to-reach consumers are more likely to carry American Express® than MasterCard®? Questions like these cannot be answered unless many more studies are done by product category.

Exhibit 6: The potential effect of non-response

	Response Rate (%)																		
Incidence among NR	95	90	85	80	75	70	65	60	55	50	45	40	35	30	25	20	15	10	5
5	31	33	34	36	38	41	43	47	50	55	61	68	76	88					
10	31	32	34	35	37	39	41	43	46	50	54	60	67	77	90				
15	31	32	33	34	35	36	38	40	42	45	48	53	58	65	75	90			
20	31	31	32	33	33	34	35	37	38	40	42	45	49	53	60	70	87		
25	30	31	31	31	32	32	33	33	34	35	36	38	39	42	45	50	58	75	
30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30	30
35	30	29	29	29	28	28	27	27	26	25	24	23	21	18	15	10	2		
40	29	29	28	28	27	26	25	23	22	20	18	15	11	7	0				
45	29	28	27	26	25	24	22	20	18	15	12	7	2						
50	29	28	26	25	23	21	19	17	14	10	6	0							
55	29	27	26	24	22	19	17	13	10	5									
60	28	27	25	23	20	17	14	10	5	0									
65	28	26	24	21	18	15	11	7	1										
70	28	26	23	20	17	13	8	3											
75	28	25	22	19	15	11	6	0											
80	27	24	21	18	13	9	3												
85	27	24	20	16	12	6	0												
90	27	23	19	15	10	4													
95	27	23	19	14	8	2													

The above table illustrates what could happen with non-response in a case in which the true incidence is 30%. Let us assume that the response rate is 50%. If the incidence rate among non-responders is 35%, then our sample could potentially show an incidence rate that is as low as 25%. (This is the value shown in the cell that intersects 50% response rate and 35% incidence rate among non-responders.)

As can be seen from the exhibit, the further away the parameter from the incidence among non-responders, the more biased our sample estimates will be. This bias increases as a function of the response rate. Conversely, the closer the parameter to the incidence among non-responders, the less the influence of non-response, no matter how severe the non-response is.

Referring once again to Exhibit 4, if a product is mass market (FMCG, general readership, ownership of standard consumer durables), the non-response is likely to follow the Case 3 pattern. If so, response rates will have minimal impact on survey estimates. But if the low response rate resembles that of Case 4 *and* if the segments that depress the response rate

Exhibit 7: Credit card ownership

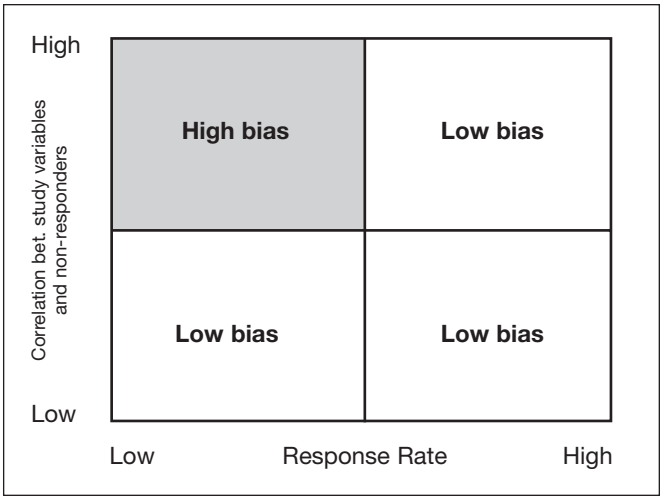
	9% RR %	31% RR %	Diff
Has Any Credit Cards	78	82	+4
Specific Cards Owned*			
Visa®	67	70	N
MasterCard®	52	48	-4
American Express®	13	18	+5
Diners®	1	1	N
Any Department Store	45	46	N
Any Gasoline Company	15	14	N
Avg. # of Cards Owned*	2.4	2.5	N
**Base: Total Credit Card owners; N = Not significant*			

Marketing Research and Intelligence Association

An advantage of focused empirical studies such as the ones we have been discussing is that, for example, if it can be established for media studies that the readership figures do not differ substantially when the response rate is 20% compared to when it is 60%, then a case can be made for lower response rates for media studies. This will substantially reduce the cost of the study, since most media readership studies spend substantial time and money increasing the response rates. Yet it should be borne in mind that a finding that is true of media studies is not necessarily true of other studies. Even if it can be established that response rate does not matter for most studies, it cannot be assumed that it would not matter for any.

Most studies carried out in the field thus far, important as they are, tend not to consider seriously the potential impact of the strength of correlation between the study variable and non-response. If we jointly consider response rates and correlation between response rate and the study variable (Exhibit 8), we see that *for three out of four quadrants, response rates may not result in high bias*. The bulk of marketing research studies are for potentially low bias mass market products such as banking, insurance, telecom, FMCG, consumer durables and the like. Consequently, the top left quadrant is likely to be less sparsely populated than the bottom right quadrant, creating a double jeopardy situation in reverse: only one of the four quadrants will potentially result in high bias, and this quadrant covers fewer products and services. From a conceptual perspective, then, we would expect that for the majority of studies, response rate would not matter.

Exhibit 8: Bias as a function of study variables correlations with non-responders



Note how only one of the four quadrants really has the potential to produce high bias due to response rate. Conceptually we would expect that, for most studies, low response rate is not likely to be a serious problem, since most marketing research studies are for products and services that are used by a large number of consumers from different backgrounds, hence are less likely to be correlated with non-responders.

What is important to note here is that results obtained using studies that are in the other three quadrants cannot be generalized to the top-left (Low Response—High Correlation) quadrant. This unintentional conceptual oversight happens mainly because the bulk of studies in applied marketing research have no discernible relationship to the study variable. While we can never rule out a correlation, there is no reason to suppose that those who respond to the survey feel substantially differently about mass market products or services compared to those who choose not to participate in surveys. When we confirm this through empirical studies, we can generalize the findings to similar product/service categories. But the results *cannot* be generalized to products/services for which there might be a strong correlation between response rate and study variables.

For a large number of products and services, responders may be similar to non-responders, so the response rate does not matter. But what about instances where they are not? This question cannot be answered unless analyses are carried out that explicitly take into account the possible correlation between the study variable and non-response.

While the online universe suffers from coverage errors (e.g., among "technophobes", less educated and less affluent consumers), the (landline) telephone universe is not immune to coverage errors, especially among younger and more mobile segments of the market—people who exclusively use mobile phones and emails to communicate—among people who register themselves in no-call lists, and among people who use caller ID as a device to talk to a select few.

Non-coverage and its effects are relatively easy to spot. The fact that online panels cannot include those who do not have access to the Internet is obvious. Non-response errors, on the other hand, can be more subtle. A researcher working with a list of those who earn over $200,000 would not be aware if a large majority of those who did not respond come from the subgroup that earns over $300,000, and therefore there might be a major difference between responders and non-responders.

There are also other pitfalls. For example, Oosterveld (2005) demonstrated how response rates can be increased in online surveys by eliminating from the panel the less-frequent responders. Eliminating low-frequency responders leaves the panel with high responders who, by definition, generate a higher response rate. High response rate achieved this way could potentially increase the bias.

We need to consider yet one more aspect before deciding on the relative merits of these two methods. This has to do with the external validity of findings based on a statistically non-representative sample.

3. Internal validity vs. external validity

The issue: *How valid are the results when we cannot guarantee that the sample is truly representative?*

In general, we need a statistically valid sample to project our results to the population. If we want to estimate the average age of Canadians, we cannot simply interview 100 people in four different Canadian cities and project the results to all Canadians. The information we collect on 100 respondents in each of the four cities does not have *external validity*, i.e., the results are not projectible to the population and we cannot know the precise quantitative relationship between the sample and the population. However, if our problem is to find out whether a given advertisement would offend women in Canada, we can conceivably ask 100 women[2] in four different cities. If a sizable proportion found the ad offensive, we can assume that, while external validity has not been conclusively proven from a statistical perspective, it can be inferred. We may not be able to quantitatively project our results to the population, but qualitatively we can assume that an ad that's offensive to a group of 100 women would be offensive to a larger group of women as well. This has been the implicit assumption in many ad testing and taste testing studies which seldom, if ever, use statistical samples. Central location ad testing and mall location taste testing have been standard methods in a researcher's arsenal for several decades. The results of such surveys have been accepted as valid over the years. (For more on internal vs. external validity refer to Liefeld 2002, 2003).

Since not all research requires external validity to be useful, the distinction between telephone interviews and online interviews may simply not be relevant for certain types of research. In such cases, what is important is which method can accomplish our objectives more effectively in terms of approach and cost.

Reformulating our concern

As we saw in the introductory section, concerns expressed about online interviewing are not that dissimilar to the ones expressed when telephone interviewing replaced personal interviewing. Criticisms were levelled against telephone interviewing: How about unlisted numbers? How about those who don't have phones? How about party lines? What if people just hang up? There were huge coverage errors with phones. Long distance charges were high; all national research firms had phone centres in major cities and conducted interviews that excluded remote areas. All such criticisms were levelled against telephone interviewing, despite the fact that door-to-door interviewing had its own limitations: it was impossible to use a truly random sample in a city when your budget was limited; "rough areas" were eliminated from the sample frame since it was not safe to go into certain areas; many apartment buildings would not allow interviewers so had to be excluded from the universe; cluster sampling, which effectively increased the margin of error, was the standard procedure, and so on. Yet today, many researchers believe that telephone interviewing is one of the best modes of interviewing. There are two reasons for this: (1) the practical effects of some early criticisms turned out to be less critical than was initially supposed, and (2) we found efficient ways to overcome some of these problems (such as using the n+ samples to include unlisted and newly listed numbers). Concerned as we are about the validity of online panels, we may yet want to reformulate our concern in more pragmatic terms.

Exploring the core issues

To summarize our discussion of the issues, coverage error and response rates share potentially the same problem: *the final sample does not resemble the universe*. The problem in the former case was created by the lack of access to certain parts of the population (coverage error) and, in the latter case, by the lack of access to certain parts of the intended sample (response rate problem). A comparison of Exhibits 1 and 4 shows that, although the sources of the problem for coverage and non-responses are different, the threats to validity within the data are similar. The problem of 88% non-response is not intrinsically "better" or "worse" than not covering 20% of the population. Both pose challenges to the validity of the results. Studies have also shown that the results of online and telephone studies can be highly correlated. How, then, do we decide between the two methods?

To really compare online panels with telephone interviews, we need to answer the following questions:

- What is the effect of coverage error on Internet surveys?
- What is the effect of non-response on both telephone and Internet surveys?

For coverage and response rates to be benign, we need to establish two things: (1) People who were not covered by the sample were not sufficiently different from those who were; and (2) People in the intended sample who did not respond to the survey are not sufficiently different from those who did.

We can also state that when the study variables are related to non-response, response rates have a much more pronounced effect than when non-response is independent of the study variables. However, since we don't know this beforehand, empirical studies such as the ones touched upon here can be very useful.

Both response rates and coverage would not matter for certain types of studies such as taste tests and advertising testing, if it can be assumed that what applies to a statistically non-representative but unbiased sample could be applied to the general population. Obviously, this will exclude studies that require certain quantitative projections such as estimating market shares.

2 Even though the sample may not be statistically random, we need to assume here that the sample is not particularly biased. For example, if all 100 women are drawn from an activist group with strong views on advertising messages, the views of this group may be at variance with that of the target population.

Reframing the issue

Like all major methodological questions, it is unlikely we can settle this issue any time soon. There are also reasons to believe that online panels are here to stay: a large and ever-increasing proportion of the population has access to the Internet, and panel members have agreed to participate in surveys. This is a tremendous advantage in a world where landlines are no longer universal and consumers can and do refuse to respond to unsolicited calls.

Given the realities on the ground, perhaps we should rethink the question. Instead of asking which method is superior, we probably should be asking which method is better for what purpose and *what the tradeoffs are*. This will probably get us out of the less productive concern of which method is superior, and move us to a more productive discussion—which method is optimal or 'satisficing' for what the researcher needs to accomplish. Then we can decide whether the disadvantages associated with the method make it unsuitable for the purpose at hand. We can start by making a conceptual list of advantages and disadvantages of each method, such as the one shown in Exhibit 9.

Exhibit 9: Relative advantages of phone and online interviews

Requirement	Relative Advantage	
Ability to demonstrate how something works		Online
Ability to locate special audience with ease		Online
Ability to persuade reluctant respondents	Phone	
Ability to show visuals to respondents		Online
Access to a better sample frame	Phone	
Adjusting to respondents' cognitive speed		Online
Completion of long questionnaires in installments		Online
Cost		Online
Coverage of certain population segments	Phone	
Coverage	Phone	
Experimental design		Online
Interviewer bias		Online
Non-intrusiveness		Online
Quick turnaround – large sample		Online
Quick turnaround – small sample	Phone	
Respondent break-off	Phone	
Respondent self-selection	Phone	
Social desirability response		Online

Conclusions

The choice between telephone interviews and online interviews should not rest on the overall superiority of one method over another. We may never be able to establish this conclusively. Neither empirical studies of the nature quoted here nor complex theoretical analyses can resolve these issues fully. Current studies seem to indicate that in many cases low-response studies, when otherwise well-designed, provide results comparable to similarly well-designed high-response studies. In many cases, well-designed online studies provide results comparable to well-designed telephone studies. An applied researcher would be better off by analyzing the problem, in addition to taking into account the results of well-controlled empirical studies, than being concerned about the overall superiority of a single method. These are some of the basic questions to consider:

(1) *Is external validity critical to this study?* If not, either of the two methods can be used. The decision would depend on cost-effectiveness and convenience.

(2) *Are quantitative estimates (such as market shares) needed?* If so, in many cases telephone studies may be better suited for the problem, at least for now.

(3) *Is coverage really the issue?* For instance, in many B2B studies telephone penetration might be the same as online penetration. In such cases the relative merits of the two would depend on the research context.

(4) *Is response rate really the issue?* As many empirical studies show, for several categories of products and services a high-response sample provides estimates that are similar to ones provided by a low-response sample. For such categories, a low response rate can be a non-issue.

(5) *Can this study really be done effectively using the other method?* If we want to do conjoint analysis, for example, it is not possible to do it using telephone surveys (unless it reduces to simple tradeoffs or the survey is sent in advance). So we have no choice with regard to the medium. Again, if we want to show a TV commercial to a wide cross-section of customers, such breadth is more easily reached online than through personal contact (such as contacting potential respondents and inviting them to a central location).

(6) *Does one or the other method create non-response related to study variables?* Response rates and coverage affect estimates only when they are related to study variables. When they aren't, they have limited impact on estimates. In general, mass market products and generalized attitudes are not affected[3] by response rates or coverage.

Empirical studies that compare low response with high response results are needed. But it should be recognized that the critical comparison is not just between high responses and low responses. It should be between high responses and low

3 This inference is based on the assumption that mass market product usage cuts across different categories of respondents. Thus, there is little reason to believe, for example, that one's usage of detergents or shampoo would be related to a person's propensity to respond to a questionnaire.

responses within a category. The critical issue is the correlation between a study variable and the response rate, not simply the response rate. Similar comments hold for coverage as well.

Acknowledgments

I would like to thank Gary Halpenny and Jeff Miller for informally sharing their views and experiences with me on the subject. The analysis presented and views expressed in this paper, however, are my own.

References

Ambrose, D. & Grey, D. (2007) Cell-only Households – A Growing concern for Telephone Surveys. *Vue*, April. Pages 20-21.

Ambrose, D., Grey, D. & Halpenny, G. (2008) Follow-up on cell-only Households – A Growing concern for Telephone Surveys. *Vue*, January, Pages 16-18.

Chakrapani, C. & Deal, K. (2005) *Marketing Research Step-by-Step*. Toronto: Pearson Education.

Deutskens E.C., Ruyter K. de, Wetzels, M.G.M. & Oosterveld, P. (2004) Response rate and response quality of Internet-based surveys: An experimental study. *Marketing Letters*, 15 (1): 21-36.

Halpenny, G. (2007) Whither Survey Response Rates: Do They Still Matter? *Staying Tuned Conference*.

Halpenny, G. & Ambrose, D. (2007) Response to Dr. Eggers. *Vue*, February 2007. Page 10.

Keeter, S., Best, J., Dimock, M. & Craighill, P. (2007) "Consequences of Reducing Telephone Survey Nonresponse". Paper presented at the annual meeting of the *American Association for Public Opinion Research*, Pointe Hilton.

Liefeld, J.P. (2002) "External (In)Validity of Characteristics of Consumer Research Reported in Academic Journals", *Canadian Journal of Marketing Research*, Vol. 20.2: 84-94.

Liefeld, J.P. (2003) Consumer Research in the Land of Oz, *Marketing Research*, Vol. 15, No. 1, Spring: 10-15.

Miller, J. (2007) Online Research Validity Discussion. *Net Gain Online Research Forum*.

Mills, D. (2007) Validity of Online Research. *Net Gain Online Research Forum*.

Oosterveld, P. (2005) Why high response rates in panels are not everything. *Research World*.

Willems, P. & Oosterveld, P. (2003) The best of both worlds, A mixture of research methods can achieve better results. *Marketing Research*, 15 (1): 23-26.

About the Author

Dr. Chuck Chakrapani is Industry Liaison Advisor and Research Mentor at the Ted Rogers School of Management, and Senior Research Fellow at the Centre for the Study of Commercial Activity (CSCA) at Ryerson University. He is also the Chief Knowledge Officer of the Blackstone Group in Chicago and President of Standard Research Systems. Chuck is Fellow of Royal Statistical Society and Fellow of the Marketing Research and Intelligence Association. He has authored more than 10 books and over 500 papers and articles. He was the former Editor-in-Chief of the *Canadian Journal of Marketing Research*. Currently, he edits *Marketing Research*, published by the American Marketing Association. Chuck consults to major firms and acts as expert witness in legal cases. He can be reached at chuck.chakrapani@gmail.com or through his website: www.ChuckChakrapani.com.

Marketing Research and Intelligence Association