

Summer 2008

Is the Digital Divide Still Closing? New Evidence Points to Skewed Online Results Absent Non-Internet Households

By Mario Callegaro, Ph.D. and Tom Wells, Ph.D.

In 2002, a Department of Commerce report titled, *A nation online: How Americans are expanding their use of the Internet*,¹ sparked a discussion around the differences between Internet and non-Internet households – the digital divide. Recently, the debate has shifted to the broadband digital divide, which shows similar findings. Now that the number of non-Internet households seems to have stabilized, we need to consider the ongoing impact of Internet penetration in the U.S. with regard to online research. Looking at figure 1, we notice a rapid trend in Internet adoption up to 2001, then a slowdown and consolidation to 64% by

Spring 2008. Data from the Pew Internet & American Life Project² shows the same type of trend in Internet penetration, with the main difference being that Pew measures Internet adoption at a “person” level, counting users who go online at least occasionally.

From these two sources of data, we can see how non-Internet households are not going to disappear anytime soon. Therefore we want to re-assess the contribution of non-Internet households to the final estimate of survey statistics, and whether we can afford to “forget about” them.

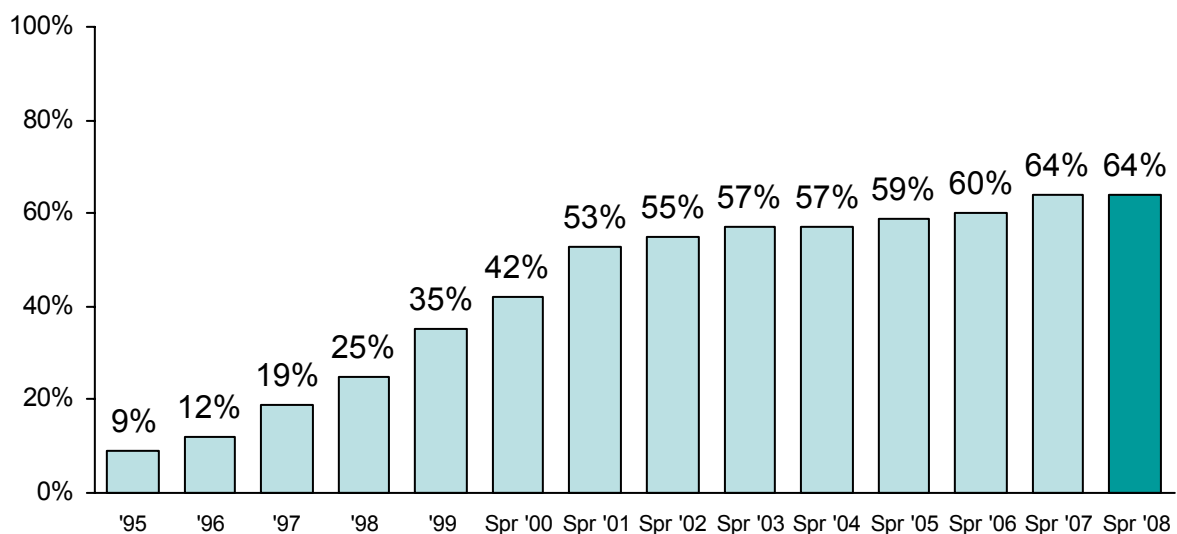


Figure 1: Historical data on Internet penetration at a household level via at-home Internet connection; Source: The Home Technology Monitor™

Impact of non-Internet households on survey estimates

Because non-Internet households have different characteristics, what is their impact on a final survey estimate? For each estimate, the impact depends on how many and how different non-Internet households are from Internet households. From a non-response point of view, the question is: *What happens if we do not talk to non-internet households?* To answer this, we present some results from late 2007 through early 2008 in table 1. It is apparent there are substantial differences between the two groups—both in attitudes and behaviors. It is worthwhile mentioning that Knowledge Networks’ (KN’s) probability-based approach³ enables us to compute confidence intervals to test whether the difference is statistically significant. The upshot: *Using an Internet-only population can produce biased results.*

Table 1. Survey estimates for selected variables by Internet, non-Internet, and total sample—by respondent %

Estimate	Non-Internet	Internet	Total	Stat Diff
Receive TV signal with a standard antenna*	26.7	16.3	21.2	Yes
Regular cable ownership*	47.0	57.8	53.8	Yes
Digital cable ownership*	51.6	40.2	44.5	Yes
Recycled your newspaper or other papers in the past 12 months*	49.1	66.7	59.6	Yes
Recycled your glass in the past 12 months*	38.2	56.4	49.1	Yes
Taken steps to reduce your use of energy in the past 12 months*	55.7	64.5	60.9	Yes
It is a citizen’s duty to keep informed about politics even if it is time-consuming**	56.8	68.1	63.5	Yes
It is a citizen’s duty to report a crime even if it might put him or her in some jeopardy**	60.8	71.1	66.9	Yes
Someone like me can’t really influence government decisions**	37.5	31.7	34.1	Yes
Do you feel that things in this country. . . have gotten off on the wrong track*	72.2	72.4	72.3	No

Note: For the measures above, one person per household was randomly selected for the analysis. In the last column we report if the difference is statistically significant in all pair-wise comparisons (Internet vs. non-Internet; Internet vs. total; non-Internet vs. total) at a .05 p level.

*: “Yes”/ “no” answer options.

** : Top two box chosen: strongly agree + agree.

The importance of non-Internet households

Leaving out non-Internet households can lead to serious over- or under-estimations. But online researchers need as clear a picture as possible of the *entire* U.S. population for:

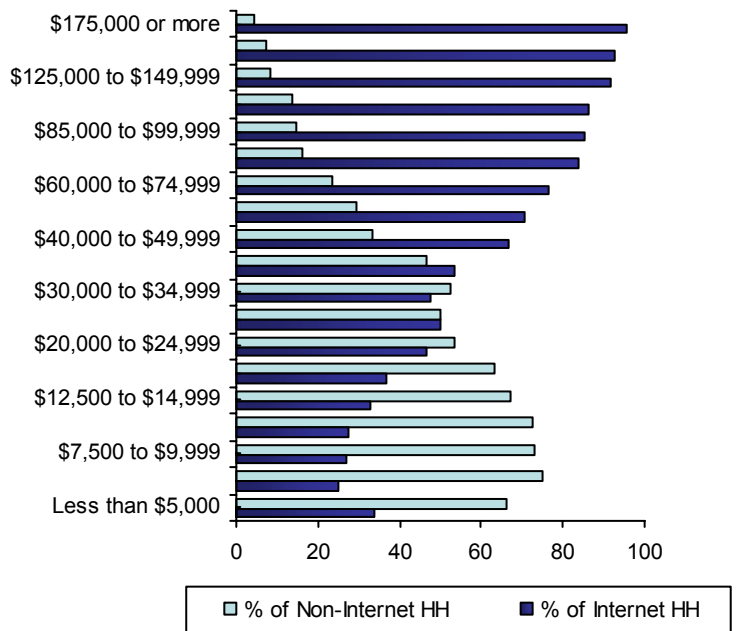
- Estimating true incidence levels
- Sizing markets and opportunities
- Obtaining publishable findings for peer-reviewed journals

Unfortunately, the largest Internet-only sample never will include this important portion of U.S. households. This factor, coupled with a multitude of additional complications—respondent self-selection; wear-out; and the ever-dwindling “reservoir” of survey-takers, to mention just a few—further negates the possibility of getting complete information. And again, it seems that non-Internet households will be around for some time.

What kinds of respondents comprise non-Internet households?

As shown in figure 2, income is the strongest predictor of being a non-Internet household, as shown via our segmentation procedure.⁴ Income and education level data share similar patterns; non-Internet households are heavily low education and low income. *Missing this sub-group can produce a distorted picture of any target audience.*

Figure 2. Household income by internet status



In the following table, we report additional characteristics of Internet and non-Internet households to provide a more complete portrait of each. A picture of non-Internet respondents emerges—most are unmarried, living in non-urban areas, and members of a minority. Our data closely follow other estimates of non-Internet status.⁵

Table 2. Selected characteristics of Internet and non-Internet households by respondent %

Ethnicity	Non-Internet	Internet
White, Non-Hispanic	30.2	69.8
Black, Non-Hispanic	60.0	40.0
Other, Non-Hispanic	26.2	73.8
Hispanic	49.1	50.9
2+ races, Non-Hispanic	39.8	60.2
Marital Status		
Married	24.9	75.1
Widowed	56.3	43.7
Divorced	49.6	50.4
Separated	59.5	40.5
Never married	45.3	54.7
Living with partner	45.4	54.6
Metropolitan Statistical Area		
Non-Urban	43.2	56.8
Urban	35.8	64.2

Note: For ethnicity and marital status, one person per household was randomly selected for the analysis. The percentage of Hispanics refers to respondents who can speak English proficiently in order to go through our recruitment call. Starting in July '08, we are including Spanish language in the recruitment call.

Can weighting correct the data?

In a previous study, authors use an RDD sample to examine attitudes on the economic outlook between Internet and non-Internet households. They show that by using model-based weighting or a more general calibration to population total, one can reduce and almost eliminate bias for these variables.⁶ A very recent paper shows similar results for health type variables, where the weighting, when applied, reduces but does not eliminate the coverage bias due to non-Internet households.⁷ However, more recent data compare results from an opt-in panel with those of a probability-based consumer panel. Even with sophisticated geo-demographic weighting, differences between Internet and non-Internet households may not be eliminated.⁸ When we conduct a preliminary analysis of the dataset in this article, using multinomial logistic regression, we see that for some variables, differences between Internet and non-Internet households still exist, even after controlling for the relevant demographic variables. This evidence provides initial proof that weighting cannot solve the problem of eliminating non-Internet households.

Conclusion

The impact of non-Internet households on survey estimates is impossible to predict in advance, and sometimes the differences are substantial. Leaving out non-Internet households can seriously lead to over- or under-estimations of the survey estimates, and most of the time it is not possible to have an external validation data to corroborate the validity of the estimates.

As marketers and academicians shift core surveys to the Internet, for accuracy's sake, it is critical to consider representation of the non-Internet population. This segment possesses a unique collective voice that is inextinguishable and central to the reliability of decisions based on online research. Those who ignore this sub-group will skew their survey results, and it is inconclusive as to whether weighting can help.

Mario Callegaro is a Knowledge Networks Survey Research Scientist. He has published nationally and internationally in the areas of telephone and cell phone surveys; polling and exit polls; longitudinal surveys; event history calendar; interviewer effect; web surveys and survey quality. He holds a B.A. in Sociology from the University of Trento, Italy, and a M.S. and a Ph.D. in Survey Research and Methodology from the University of Nebraska, Lincoln. He can be reached at mcallegaro@knowledgenetworks.com.



Tom Wells is Director of KN's Profile Group. He is responsible for designing, updating and administering our key profile surveys and overseeing our profile survey datasets. He holds a B.A. in Sociology from the University of California, Berkeley and an M.S. and Ph.D. in Sociology from the University of Wisconsin, Madison. He can be reached at twells@knowledgenetworks.com.



Footnotes and References:

- 1 Department of Commerce. 2002. A nation online: How Americans are expanding their use of internet. Washington D.C.: National Telecommunications and Information Administration.
- 2 Pew Internet & American Life Project. 2008. Percentage of U.S. households online. Retrieved June 2008, from <http://www.pewinternet.org/trends.asp>
- 3 KnowledgePanel® members are recruited using a representative probability-based sample. Non-Internet households receive both a device that connects them to the Internet and an ISP connection. Profile data in all figures and tables were collected from one member per household and weighted using KN's standard weighting procedure, which adjusts for probability of selection, demographic variables and Internet status.
- 4 The segmentation procedure used is called Automatic Interaction Detection or AID also called SEARCH procedure. The procedure starts from the entire sample cases and searches among all the possible binary splits of each predictor variable. The program then chooses the best split that maximizes the differences in distribution of the dependent variable between the two subgroups, measured by the likelihood-ratio chi-square statistics. The analysis repeats for each resulting group and the branches of the 'tree' grow until the reduction in error variance is less than 0.4 percent. More information and references are found at: http://www.isr.umich.edu/src/smp/search/search_per.html
- 5 Fox, S. 2005. Digital Divisions. Electronic Article Retrieved May 2008 Retrieved from http://www.pewinternet.org/pdfs/PIP_Digital_Divisions_Oct_5_2005.pdf
- 6 Valliant, R., & S. Lee. 2005. Economic characteristics of internet users vs. nonusers and implications for web-based surveys. *Webuse&Society*, 1, 34-51.
- 7 Dever, J. A., A. Rafferty, & R. Valliant. 2008. Internet surveys: Can statistical adjustments eliminate coverage bias? *Survey Research Methods*, 2, 47-60
- 8 Piekarski, L., M. Galin, J. Baim, M. Frankel, K. Augemberg, & S. Prince. 2008. Internet access panels and public opinion and attitude estimates. Paper presented at the 63rd Annual conference of the American Association for Public Opinion Research, New Orleans, LA.