Estimated Prevalence and Characteristics of Web Users: National Health Interview Survey, 2014-2015

Meena Khare, mkhare@cdc.gov

National Center for Health Statistics, 3311 Toledo Rd, Hyattsville, MD, 20782

Abstract[1]

Rising nonresponse and increasing survey costs continue to threaten traditional methods of data collection. In recent years, alternative survey designs (with probability or non-probability sampling) and data collection methods have been explored by survey practitioners and statistical agencies. According to the Pew Research Center telephone surveys, the rise in Internet users has grown from 14% of the U.S. adult population in 1996 to 84% in 2015. Therefore, Web surveys and multimode data collection from households and establishments have become a common, alternative cost-saving approach. The National Health Interview Survey (NHIS) has been collecting information on Internet and email use among adults since 2012. This paper presents the estimated prevalence of adult Web users (defined as an NHIS respondent who uses the Internet or email) and their sociodemographic characteristics using data from the 2014-2015 NHIS.

Keywords: Web surveys, Internet users, Email users, Noncoverage bias

1. Introduction

Rising nonresponse and increasing survey costs continue to threaten traditional methods of collecting health survey data. In recent years, alternative survey designs (with probability or non-probability sampling) and data collection methods have been explored by survey practitioners and statistical agencies. Web surveys have become a popular cost-saving alternative in marketing and other survey research. The American Association for Public Opinion Research (AAPOR) has published online reports such as "Report on Online Panels" (AAPOR, 2016, 2010; Baker et al., 2010) and "Summary Report of the AAPOR Task Force on Non-Probability Sampling" (Baker et al., 2013). In addition, a bibliography on Web survey methodology and research is available from www.WebSM.org. A few books also have been published on Web survey methodology, multimode data collection in surveys, and non-probability sampling (Callegaro et al., 2014; Couper, 2008).

According to the Pew Research Center's telephone surveys, the prevalence of Internet users has increased from 14% of the U.S. adult population in 1996 to 84% in 2015 (Keeter et al., 2015; Perrin and Duggan, 2015). Researchers from ICF International reported 86.2% of adults were Web users in 2015 (71.1% using at home, 39.8% using at work, and 66.0% using on smartphones; Boyle et al., 2015), and the American Community Survey (ACS) reported 74.4% of U.S. households as Internet users in 2013 (File and Ryan, 2014). Because of high Internet usage, Web surveys and multimode data collection from households and establishments are being explored as alternative cost-saving approaches in national surveys (Couper, 2013). However, Web surveys also experience nonresponse and noncoverage issues as with other household-based in-person or telephone surveys. Pew Research Center reported differential noncoverage by demographic and socioeconomic characteristics of adults in Web panel surveys (Keeter, et al., 2015); in 2015, the estimated prevalence of Internet users varied from a high of 96% among adults aged 18-29 years to

---

[1] The findings and conclusions in this paper are those of the author and do not necessarily represent the views of the National Center for Health Statistics, Centers for Disease Control and Prevention.

a low of 58% among older adults aged 65 and over. Pew also reported that adults with low income, less education, living in rural areas, or aged 65 and over are underrepresented among Internet users (see the Pew Research Internet Project report on the latest trends, available from: http://www.pewinternet.org/data-trend/internet-use/latest-stats/). In addition, sampling frames for volunteer opt-in Web panel surveys are generally incomplete. Moreover, variations in sample design and sampling methods between Web surveys result in different survey estimates (Keeter, 2015; Kennedy et al., 2016). Couper (2015) also discussed the role of the Web in health surveys, and issues with Web surveys, in reporting his research findings.

For this analysis, a Web user is defined as an adult NHIS respondent who identifies him- or herself as either an Internet or an email user. This paper presents the percentage of adults who reported using the Internet or emails (i.e., Web users) and compares their sociodemographic characteristics and health status or conditions using data from the National Health Interview Survey (NHIS).

2. Data and Methods

NHIS is a large health survey of the household population of the United States and covers both Internet users and nonusers. NHIS uses an area-based sampling frame of the noninstitutionalized civilian household population and a multistage sample design. A wide range of demographic, socioeconomic, and self-reported (or proxy) health information is collected using an in-person, face-to-face computer-assisted interview. In 2012, NHIS added a few questions to the adult questionnaire on use of the Internet and emails (Table 1) to evaluate feasibility of using Web surveys in future data collections.

Percentages of Web users and nonusers among 70,369 adult respondents are compared using data from the 2014-2015 NHIS (National Center for Health Statistics, 2015, 2016). Adults who do not use the Web are less likely to be included in the sampling frames for Web surveys, which could result in noncoverage of some population subgroups. Self-reported (or proxy) information on selected demographic, socioeconomic, geographic, and self-reported health conditions are compared to evaluate the characteristics of Web users and nonusers.

All data analyses, including Logistic regression models to predict the propensity (or relative odds) of being a Web user, used estimation procedures for complex surveys and the final adult analytic weights WTFA_SA to account for NHIS stratification, clustering, and nonresponse. All data analyses are done using SAS v9.3.

3. Results

Among 70,369 adult respondents aged 18 and over, the overall estimated prevalence of being a Web user was 74.1% (75.5% in 2014 and 72.7% in 2015); the estimated prevalence of Internet usage was 73.2%, which is slightly lower than other published reports. When compared by age, about one-half of adults aged 65 and over (49.3%) reported themselves as Web users. Furthermore, the estimated prevalence of Web usage was 61.3% for the age group 65-74, and it was 32.3%, the lowest, for the older age group 75 and over. This indicates that Web usage declines with age, and older adults aged 65 and over (especially those aged 75 and over) are less likely to be represented in the sampling frame for Web surveys. Because of the lower prevalence of Web usage among older adults, data for NHIS respondents aged 65 and over are excluded from this analysis, which focuses only on adults aged 18-64.

Table 2 presents the percentage of Internet, email, and Web usage among 53,347 adults aged 18-64. The overall estimated prevalence of Web usage was 79.9%, with 95% confidence limits of 79.3%—80.5%. Among adults who reported using the Internet (79.0%), 87.7% reported using email and 94.2% reported having access to a working cell phone in the family. Table 2 also compares selected demographic characteristics, socioeconomic status, and health conditions of Web users and nonusers. It shows that young adults aged 18-34 are most likely to be represented in the sampling frame for Web surveys, with 86.5% reporting themselves as a Web user. Bolded categories in Table 2 show that adults who are U.S. citizens, NH-white or NH-Asian, with greater than high school education (> HS/College and Bachelor+), with poverty-to-income ratio (PIR) > 200%, in excellent or very good health, and either married/partners or never-married singles are also more likely to be Web users. Additionally, those respondents who reported having health insurance, a working cell phone in the family, having a job in the last week or in the past 12 months, having had 12 or more alcohol drinks in any year, being a non-smoker, and reporting "No" for ever having health conditions such as diabetes, arthritis, hypertension, heart disease or heart attack, high cholesterol, or chronic obstructive pulmonary disease (data for some health conditions not shown) are more likely to be Web users.

To explore a set of significant covariates associated with Web usage (1/0), the above comprehensive list of covariates from Table 2 was included in a Logistic regression model. The forward stepwise selection method of the SAS Logistic regression with normalized sampling weights was used to identify an initial set of significant covariates, because the SAS SurveyLogistic procedure does not include stepwise methods. Next, the SAS SurveyLogistic procedure and other model diagnostics were used to select the final model, including 10 significant covariates with the predictive power of 80.4% (area under the ROC curve, c = 0.804). Table 3 presents the characteristics and the estimated relative odds of being a Web user along with 95% confidence intervals using the final Logistic regression model with 10 significant covariates. The results show that adults aged 18-34, females, NH-white or NH-Asian, having PIR > 200%, having education > HS/college or Bachelor+, U.S. citizens, having a job last week or in the past 12 months, having a working cell phone in the family, living in the West, and being in excellent, very good, or good health (self-reported or proxy) are predicted with higher relative odds of being Web users and, hence, are more likely to be represented in the sampling frames for Web surveys.

Figure 1 shows the distribution of the predicted propensities (range: 0.067—0.987) using the model with 10 covariates listed in Table 3, and Table 4 presents the basic summary measures of the predicted propensities. The median propensity indicates that 50% of the adults aged 18-64 have the predicted propensity of 0.86 or higher for being a Web user and are more likely to be represented in the sampling frames.

4. Summary and Conclusion

In 2012, a few questions were added to the NHIS adult questionnaire on Internet and email usage. Using data from the combined 2014-2015 NHIS, an estimated 79.9% of adults aged 18-64 reported themselves as being Web users (i.e., Internet or email users). Most of the adult Internet users reported using emails or having access to working cell phones in the family, which could potentially improve survey organizers' ability to contact participants in Web surveys.

The estimated prevalence of Web users declined with increasing age, and it was substantially lower among older adults aged 65 and over (49%), suggesting they are more

likely to be underrepresented in sampling frames for Web surveys. In addition, Hispanic persons, adults with lower income or PIR, less education, and with poor or fair health status also show lower representation as Web users than other subgroups.

The final Logistic regression model for predicting relative odds (or the propensity) of being a Web user among adults aged 18-64 resulted in selecting 10 significant covariates: age, sex, race and ethnicity, education, PIR, self-reported (or proxy) health status, U.S. citizenship, work status in the last week or the last 12 months, working cell phone in the family, and region (Table 3).

Findings from this assessment show differences in characteristics of Web users and nonusers and suggest that Web surveys may not be appropriate for studies targeting certain population subgroups due to possible underrepresentation in the sampling frames. For example, a Web survey may not be appropriate for studying health conditions among older adults or in populations with lower socioeconomic status. Results from Web surveys need to be carefully interpreted with respect to the survey objectives and the target population. Additional analyses may be needed to assess bias in survey estimates due to differential noncoverage or non-participation by reweighting a subsample of adult Web users and comparing the reweighted estimates with the overall NHIS estimates.

Previously published reports (AAPOR, 2016, 2010; Keeter et. al., 2015) discussed some of the issues with Web surveys presented here, such as incomplete sampling frames with noncoverage of some population subgroups (e.g., underrepresentation of adults 65 and over), and a few additional issues with nonresponse bias, selection bias, mode effect and data quality, weighting and estimation methods, and quality of survey estimates and inferences. Furthermore, additional research is needed on methods to adjust sampling weights and methods to estimate variances from Web surveys to predict uncertainties in survey estimates (Lee and Valliant, 2009; Schonlau et. al., 2009; Dever et al., 2008).

References

AAPOR (2016). AAPOR Report on "Evaluating Survey Quality in Today's Complex Environment". May, 2016. http://www.aapor.org/AAPOR_Main/media/MainSiteFiles/AAPOR_Reassessing_Survey_Methods_Report_Final.pdf

AAPOR (2010). AAPOR Report on Online Panels. Deerfield, IL: American Association for Public Opinion Research. http://poq.oxfordjournals.org/content/early/2010/10/19/poq.nfq048.full.pdf?ijkey=0w3WetMtGItMuXsandkeytype=ref

Baker, R., Blumberg, S.J., Brick, J.M., Couper, M.P., Courtright, M., Dennis, J.M., Dillman, D., Frankel, M.R., Garland, P., Groves, R.M., Kennedy, C., Krosnick, J., Lavrakas, P.J., Lee, S., Link, M., Piekarski, L., Rao, K., Thomas, R.K., and Zahs, D. (2010), "Research Synthesis: AAPOR Report on Online Panels." Public Opinion Quarterly, 74 (4): 711-781.

Baker, R.P., Brick, J.M., Bates, N.A., Battaglia, M., Couper, M.P., Dever, J.A., Gile, K.J., and Tourangeau, R., (2013), "Summary Report of the AAPOR Task Force on Non-Probability Sampling." Journal of Survey Statistics and Methodology, 1 (2): 90-143. http://www.aapor.org/AAPOR_Main/media/MainSiteFiles/NPS_TF_Report_Final_7_revised_FNL_6_22_13.pdf

Boyle, J., Fakhouri, T., Freedner-Maguire, N., and Iachan, R., (2015), "Prevalence and Characteristics of Web Panel Members." Presented at FCSM, December.

Callegaro, M., Baker, R., Bethlehem, J., Göritz, A., Krosnick, J.A., and Lavrakas, P.J. (Eds.) (2014), Online Panel Research: A Data Quality Perspective. New York: Wiley

Couper, M.P. (2008), Designing Effective Web Surveys. New York: Cambridge University Press.

Couper, M.P. (2013), "Is the Sky Falling? New Technology, Changing Media, and the Future of Surveys." Survey Research Methods, 7 (3): 145-156.

Couper, M.P. (2015, session AA6), "The Role of the Web in National Health Surveys." Presented at the 2015 National Conference on Health Statistics, Rockville, MD. August. http://www.cdc.gov/nchs/events/2015nchs/presentations.htm

Dever, J.A., Rafferty, A., and Valliant, R. (2008), "Internet Surveys: Can Statistical Adjustments Eliminate Coverage Bias?" Survey Research Methods, 2 (2): 47-62.

File, T., and Ryan, C., (2014), "Computer and Internet Use in the United States: 2013," American Community Survey Reports, ACS-28, U.S. Census Bureau, Washington, DC, 2014. https://www.census.gov/history/pdf/2013computeruse.pdf

Keeter S., (2015), "Methods can matter: Where Web surveys produce different results than phone interviews". Pew Research Center, May. http://www.pewresearch.org/fact-tank/2015/05/14/where-web-surveys-produce-different-results-than-phone-interviews/

Keeter S., McGeeney K., Mercer A., Hatley N., Patten .E, and Perrin A, (2015), "Coverage Error in Internet Surveys: Who Web-Only Surveys Miss and How That Affects Results". Pew Research Center, September. http://www.pewresearch.org/files/2015/09/2015-09-22_coverage-error-in-internet-surveys.pdf

Kennedy, C., Mercer, A., Keeter, S., Hatley, N., McGeeney,K., and Gimenez, A. (2016), "Evaluating Online Nonprobability Surveys: Vendor choice matters; widespread errors found for estimates based on blacks and Hispanics". Pew Research Center, May. http://www.pewresearch.org/files/2016/04/Nonprobability-report-May-2016-FINAL.pdf

Lee, S., and Valliant, R. (2009), "Estimation for Volunteer Panel Web Surveys Using Propensity Score Adjustment and Calibration Adjustment." Sociological Methods and Research, 37 (3): 319-343.

http://www.pewresearch.org/files/2015/09/2015-09-22_coverage-error-in-internet-surveys.pdf.

National Center for Health Statistics (NCHS). National Health Interview Survey, 2014. Public-use data file and documentation. http://www.cdc.gov/nchs/nhis/quest_data_related_1997_forward.htm. 2015.

National Center for Health Statistics (NCHS). National Health Interview Survey, 2015. Public-use data file and documentation. http://www.cdc.gov/nchs/nhis/quest_data_related_1997_forward.htm . 2016.

Perrin, A. and Duggan, M., (2015), "Americans' Internet Access: 2000-2015". Pew Research Center, June. http://www.pewinternet.org/files/2015/06/2015-06-26_internet-usage-across-demographics-discover_FINAL.pdf

Schonlau, M., van Soest, A., Kapteyn, A., and Couper, M.P. (2009), "Selection Bias in Web Surveys and the Use of Propensity Scores." Sociological Methods and Research, 37 (3): 291-318.

Appendix
Table 1: Adult Survey Questions on Internet and Email Use: NHIS, 2012-2015

Questions	Response	Estimated Prevalence* 18+ Years (%)	Estimated Prevalence* 18-64 Years (%)
AWEBUSE: Do you use the Internet?	Yes/No	73.2	79.0
AWEBORP: Research companies invite people to become members of online research panels where they regularly respond to surveys online. Are you currently a member of an online research panel?	Yes/No (added only in 2014-2015)	1.8	2.0
AWEBEML: Do you send or receive emails?	Yes/No	64.8	70.1

*One-way distribution including missing responses

Table 2: Percentage and Self-reported Sociodemographic Characteristics and Health Conditions of Internet Users, Email Users, and Web Users Among Adults Aged 18-64 Years: NHIS, 2014-2015

Characteristics*	Internet User	Email User	Web User (Internet or Email)		95% Confidence Limits Web User	
	Yes (%)	Yes (%)	Yes (%)	No (%)	Yes (%)	
Total Adults Aged 18-64	79.0	70.1	79.9	20.1	79.3	80.5
Age Group						
18-34	**85.6**	**75.5**	**86.5**	**13.5**	**85.8**	**87.3**
35-44	**80.3**	**72.0**	**81.0**	**19.0**	**79.9**	**82.0**
45-54	75.0	67.3	75.9	24.1	74.8	77.1
55-64	70.2	61.6	71.0	29.0	69.8	72.2
Sex						
Male	77.9	68.0	78.8	21.2	78.0	79.5
Female	**80.1**	**72.2**	**81.0**	**19.0**	**80.3**	**81.7**
Race/Ethnicity**						
Hispanic	65.1	52.8	66.4	33.6	65.2	67.7
NH-White	**83.8**	**75.9**	**84.5**	**15.5**	**83.8**	**85.2**
NH-Black	73.2	63.1	74.2	25.8	72.9	75.5
NH-Asian	**82.9**	**75.1**	**83.2**	**16.8**	**81.4**	**85.1**
NH-Others	69.8	61.9	71.6	28.4	65.8	77.4
Education**						
<HS	46.3	32.3	47.7	52.3	46.0	49.5
=HS	70.3	56.5	71.3	28.7	70.2	72.4
>HS/college	**86.2**	**78.0**	**87.0**	**13.0**	**86.2**	**87.7**
Bachelor+	**91.4**	**87.6**	**92.0**	**8.0**	**91.4**	**92.5**
PIR**						
<100%	62.0	49.3	62.9	37.1	61.3	64.5
100-199%	68.0	55.9	68.9	31.1	67.6	70.3
200-399%	**81.0**	**71.0**	**81.8**	**18.2**	**80.8**	**82.9**
400%+	**89.8**	**84.7**	**90.6**	**9.4**	**90.0**	**91.2**
Marital Status						
Married/partner	**79.9**	**71.4**	**80.6**	**19.4**	**79.8**	**81.4**
Widow/Divorced/Separated	70.1	61.5	71.1	28.9	69.8	72.3
Never married	**81.4**	**71.5**	**82.4**	**17.6**	**81.6**	**83.3**
U.S. Citizenship						
Yes	**81.2**	**72.7**	**82.0**	**18.0**	**81.5**	**82.6**
No	59.2	47.4	60.4	39.6	58.5	62.3
Health status						
Excellent	**85.0**	**77.3**	**86.0**	**14.0**	**85.1**	**86.8**

Very good	**83.7**	**76.1**	**84.4**	**15.6**	**83.7**	**85.2**
Good	74.7	64.0	75.7	24.3	74.6	76.8
Fair	60.0	48.5	60.9	39.1	59.1	62.8
Poor	48.4	36.0	48.9	51.1	45.3	52.5
House tenure						
Owned	**81.8**	**73.7**	**82.7**	**17.3**	**82.0**	**83.3**
Rented/other	74.7	64.5	75.4	24.6	74.6	76.3
Uninsured						
Yes	66.5	53.8	67.6	32.4	66.1	69.1
No	**81.1**	**72.9**	**81.9**	**18.1**	**81.4**	**82.5**
Working cell phone in family						
Yes	**80.0**	**71.2**	**80.9**	**19.1**	80.3	81.4
No	66.2	55.4	66.9	33.1	64.9	69.0
Region						
Northeast	80.0	71.3	80.8	19.2	79.5	82.1
Midwest	80.5	71.1	81.2	18.8	80.0	82.4
South	76.6	67.2	77.5	22.5	76.4	78.6
West	**80.8**	**72.9**	**81.7**	**18.3**	**80.5**	**82.8**
Work Status						
Had job last week	**82.6**	**74.7**	**83.4**	**16.6**	**82.8**	**84.0**
No job last week, had job past 12 months	**82.8**	**72.9**	**83.7**	**16.3**	**82.1**	**85.4**
No job last week, no job past 12 months	66.9	55.6	67.7	32.3	66.3	69.1
Never worked	63.9	50.2	65.0	35.0	62.4	67.6
Ever had 12+ drinks in any one year						
Yes	**82.9**	**75.2**	**83.7**	**16.3**	**83.0**	**84.3**
No	73.1	61.7	74.0	26.0	73.1	74.9
Ever smoked 100 cigarettes						
Yes	76.1	65.8	76.9	23.1	76.0	77.9
No	**81.0**	**72.8**	**81.8**	**18.2**	**81.2**	**82.5**
Ever been told that you have diabetes						
Yes	64.0	54.3	65.1	34.9	63.0	67.1
No	**80.2**	**71.4**	**81.0**	**19.0**	**80.4**	**81.6**
Pre-diabetic	74.4	62.5	74.7	25.3	70.3	79.1
Ever been told you had arthritis						
Yes	73.5	64.3	74.1	25.9	72.7	75.5
No	**80.1**	**71.3**	**81.0**	**19.0**	**80.4**	**81.6**
Ever been told you have hypertension						
Yes	72.6	63.5	73.5	26.5	72.4	74.6
No	**81.1**	**72.2**	**81.9**	**18.1**	**81.3**	**82.5**

 *Bold: Categories with high percentages**;**
 **NH (Non-Hispanic); HS (High School); PIR (poverty-to-income ratio or poverty status);

Table 3: Estimated Odds Ratio of being a Web User Among Adults 18-64 Years Using a Logistic Regression Model: NHIS, 2014-2015

Effect Adults 18-64 years	Estimated* Odds Ratio	95% Confidence Limits	
Age Group 35-44 vs 18-34	0.554	0.503	0.610
45-54 vs 18-34	0.375	0.341	0.413
55-64 vs 18-34	0.270	0.244	0.298
Sex Male vs Female	0.817	0.767	0.870
Race/Ethnicity** NH-White vs Hispanic	1.508	1.374	1.656
NH-Black vs Hispanic	1.119	1.007	1.242
NH-Asian vs Hispanic	1.271	1.079	1.499
NH-Others vs Hispanic	0.980	0.711	1.351
PIR** 100-199% vs <100%	1.213	1.104	1.332
200-399% vs <100%	1.813	1.643	2.001
400%+ vs <100%	2.905	2.605	3.240
Education** =HS vs <HS	1.878	1.710	2.063
>HS college vs <HS	4.223	3.824	4.663
Bachelor+ vs <HS	5.772	5.153	6.464
Health Status Very good vs 1 Excellent	1.029	0.934	1.133
Good vs 1 Excellent	0.883	0.800	0.975
Fair vs 1 Excellent	0.673	0.600	0.755
Poor vs 1 Excellent	0.496	0.406	0.607
U.S. Citizenship No vs Yes	0.608	0.546	0.677
Work Status No job last week, had job past 12 months vs Had job last week	1.293	1.130	1.480
No job last week, no job past 12 months vs Had job last week	0.873	0.799	0.954
Never worked vs Had job last week	0.693	0.601	0.799
Working cell phone in Family: No vs Yes	0.692	0.621	0.772
Region Midwest vs Northeast	0.977	0.855	1.116
South vs Northeast	0.929	0.817	1.056
West vs Northeast	1.296	1.141	1.472

**NH (Non-Hispanic); HS (High School); PIR (poverty-to-income ratio or poverty status);

Table 4: Basic Statistical Measures of Propensity of being a Web User, Using Logistic Regression Model Covariates in Table 3: NHIS, 2014-2015

Basic Statistical Measures of Propensity Distribution			
Minimum	0.067067	Maximum	0.986700
Mean	0.785770	Median	0.864323
Quartile 1	0.699538	Quartile 3	0.928323
Mode	0.970288	Standard Deviation	0.19417
Range	0.91963	Variance	0.03770

Distribution and Probability Plot for Propensity
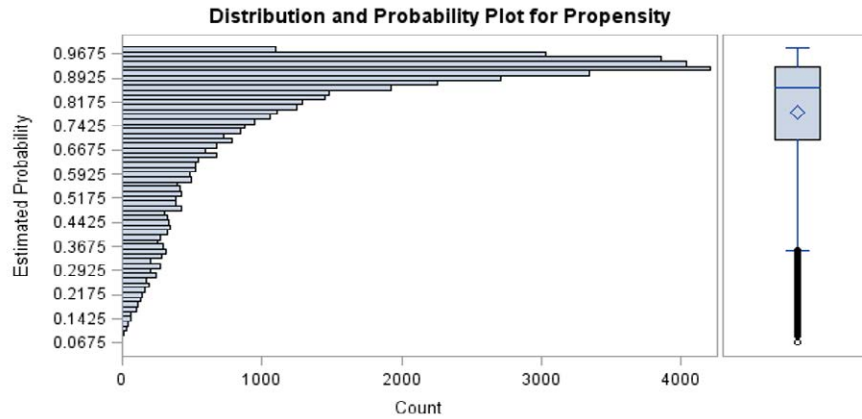
Figure 1: Distribution of the Predicted Propensity of being a Web User, Using Logistic Regression Model Covariates in Table 3: NHIS, 2014-2015