

Response to OMB's "Initial Review" 12/8/2009

1. *Questions 1-18 seem to generate enough market information for the intent of this survey without the hypothetical's that follow them. Why are the hypothetical questions needed?*

Even though questions 1 - 18 provide useful market information, there is insufficient variation in the levels of the product features to estimate important marginal utility parameters of interest. For example, monthly cost and speed of service are highly collinear. A designed choice experiment with hypothetical questions avoids this problem by manipulating the levels of the features to obtain optimal variation. In addition, we have econometric methods that combine the market information from questions 1 - 18 with the hypothetical data in a well-specified discrete-choice model to obtain very precise estimates of consumer preferences.

In addition, the answers to the hypothetical questions allow us to estimate the demand for features and bundles of features that are not currently traded (or only available in limited geographical areas) in real markets. For examples, the mobile laptop feature is not widely available; the health feature is not currently bundled into Internet service.

2. *Assuming that the hypothetical questions are necessary, we are concerned about the number of question. There might be significant survey fatigue on the part of respondents.*

This is an really important point that we have given much thought. Carson et al (1994) review a range of choice experiments and find respondents are typically asked to evaluate from one to 16 choice questions, with the average being around eight questions per respondent. Brazell and Louviere (1997) show equivalent survey response rates and parameter estimates when they compare respondents answering 12, 24, 48 and 96 choice questions in a particular choice task. In our own research (Savage and Waldman, 2008) we found there is some fatigue in answering eight choice questions when compared to mail respondents. To remedy this, we have reduced the cognitive burden to the respondent in this survey in two ways:

- by decreasing the number of features to be compared from five to four; and

- splitting the eight choice questions into two groups with a different fourth activity feature: questions 21, 23, 25, and 27, and questions 30, 32, 34, and 36. This gives the respondent a break from the choice task with an open-ended valuation question.

3. *The questions on speed seem vague. How do we know "fast" to one person means the same as another? Why not use more defined speed tiers as was done in the consumer survey?*

In our focus groups, both now and for our previous research (Savage and Waldman, 2008), subjects appeared comfortable with our three tiers and the brief characterization of speed. Moreover, they resemble descriptions used by Internet service companies in their marketing campaigns.

But for those that are not entirely sure what "fast" means, for example, we provide additional, more objective description in a hyperlink:

"Speed: This is the time it takes to receive (download) and send (upload) information from your home computer. Speed can be slow (similar to travelling on a San Francisco cable car at 5 mph), fast (similar to travelling on an AMTRAK train at 100 mph, or, 20x faster than Slow) or very fast (similar to travelling on the "bullet train" at 300 mph or, 60x faster than Slow)."

We really like the advantage of an Internet survey here: only those unsure of Speed will click on the hyperlink and take the time to read the enhanced description, thus reducing potential survey fatigue.

4. *Why is there a health medical record question being asked on a willingness to pay survey?*

We are trying to value certain possible Internet activities that have the potential to improve societal welfare, through improved communication and reduced transport costs. Online access to one's health records and doctors' notes is one such activity. This has been raised in the health and communications literatures, and in discussions with the members of the Broadband Task

Force Initiative. By including "health" in our hypothetical Internet service options, we can "back out" consumer valuation for online health services in a cost effective way.

5. *This survey mentions its being used by the University of Colorado. Any surveys that are granted OMB approval require that the final instrument that will actually be used, be uploaded for the review. So if there is a later one please send it.*

The version sent to the OMB on Thursday, December 3 is almost precise in language and is full and complete in content. Currently, the programmers at KnowledgeNetworks.com are refining the survey according to their templates and protocols. We will send the latest version of the survey to the OMB.

6. *The methods for maximizing response rates in section B.3 need to refer specifically what will be done in the current study, not what could be done.*

For our study, Knowledge Networks (KN) will

- send up to 3 email reminders to respondents who have not completed the survey in a timely manner. In KN's experience, Email reminders have proven to achieve within-survey cooperation rates of 65% and greater.
- KN allows the respondent a two to four week period to reply. This flexibility also improves response rates.
- Use our University name in the email invitation.
- Place telephone reminder calls to nonresponders.

7. *If FCC intends to use results from the Knowledge Panel to produce estimates that are nationally representative, clear non-response bias analysis plans must be provided.*

The Knowledge Panel contains more than 2,500 variables on each member, including a comprehensive set of demographics. These data enable us to know a great deal about those panelists that choose not to respond to any specific survey and these data form the core of the information used in creating non-response bias analyses. The key element in determining bias is whether or not the responders are very different from the non-responders on some important characteristic. These data make such detailed analyses possible.

If an analysis of the demographics of responders and nonresponders warrants, we will rigorously estimate model parameters corrected for non-response bias, along the lines of the econometric sample selection literature (see Heckman, 1979; Ozog and Waldman, 1996; Waldman, 1981). The steps for this analysis are:

- Fit a probit model of responders/nonresponders as a function demographic variables.
- From the estimates of the probit model, construct a Heckman-type correction term (inverse Mill's ratio);
- Add the Heckman correction in the specification of the second stage regression of the utility function.

*8(a). Please provide more information on the design of the study, specifically, is it a between-subjects or within-subjects design, what factors are being manipulated and how many different conditions are there.*

The design is between subjects. There are four factors (features) that are manipulated: cost, speed, reliability, and Internet activity (one of the five: Priority; Health; Mobile Laptop; Videophone; Movie Rental). This results in five survey versions:

- 1) Priority-Health;
- 2) Health-Mobile laptop;
- 3) Mobile Laptop-Videophone;
- 4) Videophone-Movie rental;

5) Movie rental-Priority.

For example, the Health-Mobile Laptop version contains four choice questions with Mobile Laptop as the Internet activity, and then four questions with Health as the Internet activity.

The description of these factors is:

**Cost:** This is how much your household pays per month for the Internet service at your home.

**Speed:** Speed of receiving and sending information over the Internet

**Slow:** Similar to dial up. Downloads from the Internet and uploads to the Internet are slow. It is good for emailing and light web surfing.

**Fast:** Much faster downloads and uploads. It is great for music, photo sharing and watching some videos.

**Very Fast:** Blazing fast downloads and uploads. It is really great for gaming, watching high-definition movies, and instantly transferring large files.

**Reliability:** Very reliable Internet service is rarely disrupted by service outages, that is, your service may go down once or twice a year due to severe weather. With less reliable Internet service you will experience more outages, perhaps once or twice a month for no particular reason.

**Mobile laptop:** This feature allows you to use your Internet service to connect your laptop to the Internet wirelessly while away from your home.

**Health:** This feature allows you to use your Internet service to access your medical records over the Internet and interact with your health care providers, saving you a trip to your Doctor, specialist or pharmacy.

**Movie rental:** This feature allows you to use your Internet service to regularly download high definition movies and TV shows from the Internet, and watch them on your computer or TV (saving the cost of a trip to the video store)."

Videophone: This feature allows you to use your Internet service to place free phone calls over the Internet and see the person you are calling.

Priority: This feature allows you to designate some of your Internet downloads as high priority. During peak times, your high-priority downloads will travel through the Internet at a much faster speed than low-priority downloads.

The levels of these features are:

Cost: \$5 to \$90/month (\$5 increments)

Speed: Very fast, Fast, Slow

Reliability: Very reliable, Less reliable

Internet Activity: Yes, No

Measures developed by Zwerina et al (1996) generate an efficient non-linear optimal design. A fractional factorial design creates 24 paired descriptions of Internet service that are grouped into three sets (1, 2, and 3) of eight questions, randomly distributed across all respondents. See the table below.

3	2	B	Very	No	Fast	\$85.00
3	3	A	Less	No	Slow	\$30.00
3	3	B	Less	No	Very Fast	\$55.00
3	4	A	Less	No	Slow	\$25.00
3	4	B	Less	Yes	Slow	\$30.00
3	5	A	Very	No	Fast	\$75.00
3	5	A	Very	Yes	Slow	\$5.00
3	6	B	Less	No	Very Fast	\$65.00
3	6	B	Less	No	Very Fast	\$35.00
3	7	A	Less	No	Fast	\$55.00
3	7	B	Very	No	Very Fast	\$80.00
3	8	A	Less	Yes	Fast	\$10.00
3	8	B	Very	Yes	Slow	\$20.00

In summary, the five survey versions described above and the three choice sets from the fractional factorial design result in a final set of 15 survey versions:

- 1) Priority-Health 1;

- 2) Priority-Health 2;
- 3) Priority-Health 3;
- 4) Health-Mobile laptop 1;
- 5) Health-Mobile laptop 2;
- 6) Health-Mobile laptop 3;
- 7) Mobile Laptop-Videophone 1;
- 8) Mobile Laptop-Videophone 2;
- 9) Mobile Laptop-Videophone 3;
- 10) Videophone-Movie rental 1;
- 11) Videophone-Movie rental 2;
- 12) Videophone-Movie rental 3;
- 13) Movie rental-Priority 1;
- 14) Movie rental-Priority 2;
- 15) Movie rental-Priority 3.

These will be randomly assigned across the 4,500 respondents, resulting in approximately 300 per version. These versions are identical except for the Internet activity being evaluated, and the levels of all features.

*8b) If there are any within-subjects factors, explain what is being done to counter-balance potential order effects.*

There are no within-subject factors.

8c) *Also provide analysis plans and a clear justification for the sample size needed given the analytic goals.*

The analysis plan is to estimate the parameters of individual's utility function, and then to use these estimates to construct estimates of the willingness-to-pay (WTP) for features of an Internet service. The reason for the choice questions is that estimated WTP from open-ended questions are unreliable: individuals tend to overestimate their values when they do not face a clear comparison. Answers to such questions require maximum likelihood estimation. Maximum likelihood estimates have desirable asymptotic properties, which require large samples to take effect. While estimating a population mean from a normal population with a random sample may require only 50 or 100 observations, our likelihoods are nonlinear and extremely complex, requiring many more observations. In addition, we are trying to value five Internet activities, but feel we cannot ask five Internet activity questions to each respondent. Therefore, we have split these questions up, as described above. The effective sample size for each activity is therefore  $4500/5 = 900$ . In addition, we will focus on particular subsamples of the population:

- those with and without Broadband;
- high income vs. low income;
- highly educated vs. less so;
- rural vs. urban residents.

The sample of 4,500 observations will accommodate the precise estimation for each activity and subsample.

Response to OMB's "Subsequent Review" 12/11/2009

1. *Provide a structural economic and econometric model*

For a detailed discussion of the economic model of the demand for Internet access and the random utility econometric model, please see Appendix A: *Structural economic and econometric model*.

2. *Intuitive explanation of the specific parameters of interest to be estimated from the structural model: marginal utilities and willingness-to-pay*

The estimated parameters of interest are the marginal disutility of cost and the marginal utilities for all the features that comprise Internet service, i.e., the  $\beta$ 's in the utility function (U) below:

$$U^* = \beta_1 \text{COST} + \beta_2 \text{SPEED} + \beta_3 \text{RELIABLE} + \beta_4 \text{MOBILE LAPTOP} \\ + \beta_5 \text{MOVIE RENTAL} + \beta_6 \text{VIDEOHONE} + \beta_7 \text{PRIORITY} + \beta_8 \text{HEALTH} + \varepsilon \quad (1)$$

The marginal utilities have the usual partial derivative interpretation: the change in utility from a one unit increase in the level of the feature. Given vertical differentiation, we would expect, for example, that more speed would provide more utility so that  $\beta_2 > 0$ . For example, an estimate of  $\beta_2 = 0.1$  indicates that a one unit improvement in SPEED (e.g., the discrete improvement from "Slow" to "Fast") increases utility by 0.1 for the representative consumer. In contrast, a higher cost of service provides less utility (or, disutility), so  $\beta_1 < 0$ .

Since the estimates of marginal utility for the features do not have a readily understandable metric, it is convenient to convert these changes into dollar terms. This is done by employing the economic construct called willingness-to-pay (WTP). For example, the WTP for a one unit increase in SPEED (e.g., the discrete improvement from "Slow" to "Fast") can be interpreted as how much more the Internet service would have to be priced to make the consumer just indifferent between the old (cheaper but slower) service and the new (more expensive but faster) service. The required change in cost to offset an increase of  $\beta_2$  in utility is, from equation (1):

$$-\beta_2/\beta_1 \quad (2)$$

For example, estimates of  $\beta_2 = 0.1$  and  $\beta_1 = -0.01$  indicate that the WTP for a improvement in speed from “Slow” to “Fast” is \$10 ( $= -0.1/0.01$ ). Note that the model specification (1) implies that the representative consumer would also be willing to pay \$10 for an improvement in speed from “Fast” to “Very Fast.” This constraint can be easily relaxed during econometric estimation so that the WTP for an improvement in speed from “Fast” to “Very Fast” can be different to the WTP for an improvement in speed from “Slow” to “Fast.”

This approach to estimating consumer valuations is true for all other features and Internet activities. The WTP for RELIABLE, MOBILE LAPTOP, MOVIE RENTAL, VIDEOPHONE, PRIORITY and HEALTH is the ratio of its marginal utility to the marginal disutility of COST. In summary, the WTP construct provides a theory-driven but very intuitive (dollar) measure of the value consumers place on Internet service, and the specific features and activities that comprise the service.

Since the WTP estimates are nonlinear functions of the structural parameter estimates, estimation of their standard errors for the purpose of hypothesis testing is complex. See Appendix B: *Estimating the standard error of WTP measures from discrete choice experiments.*

Individuals may not have identical preferences. An individual’s preference toward speed, for example, may differ because of observable demographic characteristics, or may be idiosyncratic. It is possible to estimate differences in the marginal utility of specific service features to different consumers by interacting those features with demographic variables. For instance, suppose individuals in different locations (rural vs. urban) value speed differently. A model that captures this difference is:

$$U^* = \beta_1 \text{COST} + (\beta_2 + \eta \text{RURAL}) \times \text{SPEED} + \beta_3 \text{RELIABLE} + \beta_4 \text{MOBILE LAPTOP} \\ + \beta_5 \text{MOVIE RENTAL} + \beta_6 \text{VIDEOPHONE} + \beta_7 \text{PRIORITY} + \beta_8 \text{HEALTH} + \varepsilon \quad (3)$$

where  $\eta$  is an additional parameter to be estimated, and RURAL is a dummy variable that is equal to one when the respondent is in a rural location, and zero otherwise. When rural location is not important ( $\eta = 0$ ), the WTP for a one-unit improvement in speed is  $-\beta_2/\beta_1$ . When rural location is important ( $\eta \neq 0$ ), the WTP for a one-unit improvement in speed in a rural location is:

$$\frac{(\beta_2 + \eta RURAL)}{\beta_1} \quad (4)$$

Equation (4) provides a concrete illustration of how the WTP estimates from this study will inform the design of government programs to promote Broadband Internet access in under-served areas. For example, policy makers can use (4) to compare rural valuations for Broadband to the cost of service provision, and then make a more accurate judgment of the potential subsidy required to incent individual Broadband adoption and/or deployment in rural areas.

### 3. *Explain your plan to address non-response bias*

In addition to our discussion in point 7 of the “Response to OMB’s Initial Review of 12/8/2009,” Dennis (2009) describes the statistical techniques used in the KN within-panel sampling methodology and, importantly, provides evidence of sample representativeness of the general population. For evidence of the representativeness of KN panel samples, please refer to Dennis (2009; pp. 3-5), available at:

<http://www.knowledgenetworks.com/ganp/docs/KN-Within-Panel-Survey-Sampling-Methodology.pdf>

Moreover, a list of selected OMB approved research projects where KN was the data collection subcontractor is available at:

<http://www.knowledgenetworks.com/ganp/docs/OMB-Project-Approvals-for-KN.pdf>

However, to be confident in our sampling scheme, it may be important to account for any systematic variation in important demographic variables between the population and our KN sample. We will test the equality of means of demographic variables, such as age, race, gender, education and income, between our data and U.S. Census Bureau data. If we find that our sample differs significantly from the Census we will apply post-stratification weights to our KN sample and use weighted, maximum likelihood estimation to estimate our econometric model. See Savage and Waldman (2008) for a description of this approach.

4a. *Additional details on study design: between vs. within subjects*

We misunderstood this question in the initial review. The design uses variation both between and within subjects. The within variation comes from the eight repeated A vs. B choice questions plus the follow up status quo question (i.e., the choice between current home service and choice A or B). The analysis must take into consideration the fact that the formation of that part of the likelihood involving the comparison of the chosen alternative to the status quo involves the error difference, and from choice occasion to choice occasion these error differences are correlated. This correlation is induced by the common occurrence of the status quo error, since respondents evaluate their utility of the status quo only once. In this study, we treat the person, and not the person-choice occasion, as the unit of observation, so that we may explicitly model this correlation. See Appendix C: *Details on the study design: within subjects*.

4b. *Additional details on study design: potential order effects*

To account for possible order effects in the eight A-B choices questions, KN will randomly assign the ordering of the eight A-B choice questions within each choice set (1, 2 & 3) across all respondents.

5. *Targeting those new to the Internet*

Based on their recruitment information, KN know if a household previously had Internet access, and the type of access (dial-up, cable modem, DSL, etc.). We will use this information to oversample new recruits to the panel (i.e., panel members with less than 12 months of panel experience) *and* who did not have Internet access prior to recruitment. This will provide a subsample of “recently connected” dial-up users to approximate households that are not connected to the Internet. There are 915 panel members that fulfill this criteria. Given a 65 percent response rate, this would provide a subsample of 595 “recently connected” dial-up users.

These data will be used to estimate the marginal utilities and WTP for various sub-samples of the population:

- new recruits with a dial-up connection and no Internet access prior to joining the KN panel;
- respondents with traditional dial-up access through a telephone modem; and
- respondents with high-speed access.

## 6. *Health as an Internet Activity*

As part of National Broadband Plan, the Broadband.gov Taskforce lists “Telehealth and Telemedicine” as one of the advantages of Broadband:

“Broadband can facilitate provision of medical care to unserved areas and underserved populations through remote diagnosis, treatment, monitoring, and consultations with specialists ([http://www.broadband.gov/broadband\\_advantages.html](http://www.broadband.gov/broadband_advantages.html).)”

Our economic modeling shows that activities, such as Telehealth and Telemedicine, have enormous potential to improve societal welfare, through improved communication and reduced transport costs (See Appendix A). As mentioned above, by including “health” in our hypothetical Internet service options, we can “back out” consumer valuation for online health services in a cost effective way (i.e., it is *not* necessary to design separate health plan choice experiments where consumers choose between different health plans with and without an online health feature). We have changed “health” to “telehealth” and re-written the description (see below) to provide a cleaner and more direct representation of this feature:

Telehealth: This feature allows you to use your Internet service for remote diagnosis, treatment, monitoring and consultations, saving you a trip to your health specialists.

Because we know the geographical location of respondents, and the deployment of Broadband, we will use the WTP construct described in equation (4) to estimate consumer valuations for Telehealth in remote and underserved locations.

One of the advantages of our survey methodology and experimental design is that we have the flexibility to add or omit Internet service features and activities without affecting the estimation

of the remaining features. For example, telehealth could be omitted from the study without any impact on the valuation of other features and activities.

### References

Brazell, J., and Louviere, J., 1997. Respondent's Help, Learning and Fatigue. Presented at the 1997 INFORMS Marketing Science Conference, University of California, Berkeley.

Carson, R., Mitchell, R., Haneman, W., Kopp, R., Presser, S., and Ruud, P., 1994. Contingent Valuation and Lost Passive Use: Damages from the Exxon Valdez. Resources for the future discussion paper, Washington, D.C.

Dennis, Michael, 2009. Description of Within-Panel Survey Sampling Methodology: The Knowledge Networks Approach, Government and Academic Research, Knowledge Networks.

Heckman, James J., Sample selection bias as a specification error, 1979, *Econometrica*, 47, 153-161.

Ozog, M., and D. Waldman, 1996. "Voluntary and Incentive-Induced Conservation in Energy Management Programs", *Southern Economic Journal*, 62, 1054-71.

Savage, Scott, and D. Waldman, 2008. "Learning and Fatigue During Choice Experiments: A Comparison of Online and Mail Survey Modes," *Journal of Applied Econometrics*, 23, 351-371.

Waldman, D., 1981. "One Economic Interpretation of the Parameter Constraints in a Simultaneous Equations Model with Limited Dependent Variables," *International Economic Review*, 22, 731-739.

Zwerina, K., Huber, J., and Kuhfeld, W., 1996. A General Method for Constructing Efficient Choice Designs. in *Marketing Research Methods in the SAS System*, 2002, Version 8 edition, SAS Institute, Cary, North Carolina.