Institute of Museum and Library Services (IMLS)
The Impact of Free Access to Computers and the Internet in Public Libraries

B. Collection of information employing statistical methods

The following offers a detailed explanation of the statistical methodology of data collection and analysis for a national telephone survey and an online, web-based survey. All methods adhere to the Office of Management and Budget's <u>Standards and Guidelines for Statistical Surveys</u> (September 2006).

The triangulated, mixed-method approach of the proposed study will yield statistically generalizeable data from people nationwide who represent socio-demographically diverse sections of the population, as well as case studies of best practice libraries in which rich, story-driven data will be obtained from key stakeholders to provide deeper insights into possible causes for the results found in the statistical analysis.

B1. Respondent universe and sampling methods

Generally, the respondent universe represents persons in the U.S. who are aged 14 and over and have used public access computers in U.S. public libraries at least once in the past year (from the day of the survey screening). Quantitative data will be collected through a nationwide telephone survey and a web survey accessed through public library websites. In addition 5 public libraries will be selected for case studies.

B1.1 Telephone survey

Telephone Contact, Inc. (TCI), will conduct a national telephone survey of public access computer users with an over sampling of low income respondents.

The target population will be persons age 14 or older who have used public access computers (PAC) in public libraries or library websites in the past year. The goal will be to contact a random sample of households by telephone in order to find 760 PAC users who are age 14 or older and willing to complete the survey. This will include some oversampling of low income telephone exchanges so that at least half of the interviews come from respondents whose household income is less than or equal to 200% of the federal poverty level.

We will use a dual frame approach with the first sample frame being a list assisted random digit dial (RDD) sample with a geographic oversample of telephone exchanges that primarily service low income neighborhoods. The goal is to complete 600 interviews using the RDD sample frame. Since the RDD frame captures a negligible fraction of cell-phone-only households, a second sampling frame will utilized: a cell phone exchange sampling frame. The goal is to complete 80 interviews using the cell phone sample frame. The frames will be stratified by geography (state and urbanicity) as well as income status (with a modest oversample of low income strata).

In addition, the telephone survey will include a non-response study described in section B2.1 that will yield an additional 80 interviews for an overall total of 760 PAC users.

Households will be screened to establish the eligibility of residents aged 14 and over. One eligible person will be randomly selected (using the 'last birthday' method) for interviewing when two or more residents are eligible within a household. Qualified respondents are those who in the past year have either used a public access computer, are aged 14 or older, and are residents of the household. By virtue of the screening protocol, limited information will be collected from persons who do not use public access computers.

Finally, we note that for the *cell phone* sample, we will not screen out cell phones that are associated with a landline household. We do this for two reasons. First, there is evidence that some household (labeled "mostly-wireless" in the survey research community) with both cell phone and landline phones rarely or never use their land line phones for incoming calls (e.g., landline is reserved for DSL service, or fax, or at-home business). Secondly, the population we have targeted is sufficiently rare that it would be inefficient from both cost and statistical perspectives to screen out a household that would otherwise be substantively "eligible." Questions in the survey instrument will allow us to separate the cell-phone-only households from those with landlines so that weighting can be developed separately for these groups.

B1.2 Web survey

The primary difficulty with relying on the telephone survey alone to collect information about public access computer users is the low prevalence of the target population and the difficulty in gauging the public access computer resources available to respondents. Moreover, we expect a small to modest proportion of PAC users to be homeless or otherwise not residing in a 'household'. To assess the impact of resource availability on public access computer use and users and to generalize findings at the community and library levels, we will draw a random sample of library systems to participate in a web survey. The web survey will also enable us to supplement the telephone survey data by reaching low income or homeless persons who may not own telephones and a greater number of respondents in the 14-17 age range.

The target population for the web survey will be persons using public access computing in public libraries or through library websites during the data collection period who are aged 14 or over. A two stage sample will be employed:

- The first stage of selection will involve a stratified probability sample of 500 library systems[1] with probability proportional to size using an appropriate measure of size such as the number of PAC stations or the population served by the library;
- At second stage of selection, all PAC users within all libraries in the sampled library systems will be asked to participate over a specific set of consecutive field days with the goal of an average of 200 completed interviews per sampled library system.

---

[1] Within the largest library systems such as New York and Los Angeles, "PSU equivalents" (representing a collection of libraries exhibiting a minimum combined measure of size) will be formed and sampled in order to reduce the burden on these large systems.

Ignoring the impact of nonresponse and multiple users, if the measure of size is highly correlated to the number of PAC users (as one would reasonably expect), then the two-stage design produces an approximately equal probability sample (epsem) of PAC sessions. However, individuals' frequency of PAC station usage and unit nonresponse prevent us from treating the resulting sample as an epsem design. We will collect individual frequency of usage and daily numbers of users to reflect differential selection probabilities in the sampling weight. Moreover, the consecutive data collection segments will be spread across days of the week and times of the day to ensure that the full diversity of usage is being captured in the sample.

We will select 500 library systems (expecting at least 400 to participate) that are representative in terms of size distribution with an oversampling of library systems that primarily serve neighborhoods associated with low income census tracts. In order to select the library systems we will need a measure of size and location which will be obtained from the already available 2006 National Center for Education Statistics (NCES) national survey. Once the library systems are selected the goal will be to complete on average about 200 interviews per library system. The questionnaire will include questions about respondents' households including information about their phone service in order to help verify telephone survey results. A primary objective of the web survey will be to explore the effect library policies and resources have on PAC use.

Within the library systems, a random sample of public library administrative units will be drawn from the NCES database. Selected libraries will be asked to post a link to the web survey on their public access computer home pages. The Chief Officers of State Library Agencies (COSLA) association will assist in coordinating the web survey through state library agencies. We expect approximately 80% of selected public library administrative units to participate by linking to the survey on their library web sites and public access computer homepages and approximately 12.5% of public access computer users to complete the survey.

B1.3 Case studies

The selection of sites for the case studies will be based on available information about public libraries and feedback from COSLA officers and state librarians. The sites will be selected to illustrate a range of public access computer systems and users. While there are too few case studies to provide a statistically or nationally representative sample, we will strive for a sample that gives a picture of the full range of public access computer users; for example, we will look for libraries that illustrate differing sizes, operate in urban or rural settings, and serve minority populations. The case studies will be used to provide a more complete picture of the impact of public access computers. In each site, we will select users, librarians, and systems administrators, as well as staff from agencies that refer people to libraries for public access computer use.

B2. Procedures for the collection of information

B2.1 Telephone survey

Telephone Contact, Inc. will conduct the interviews and program the questionnaire into a CATI (Computer assisted telephone interviewing) program. A CATI program will help interviewers

with skip patterns, call scheduling, and tracking, and has many other important advantages over paper form administration.

Under our current design parameters and response rate goals, we will release into the field roughly 57,400 telephone numbers for the RDD sample and about 7,600 telephone numbers will be needed for the cell phone sample. Tables 1 and 2 below present the expected dispositions of the RDD and Cell-Phone Telephone samples, respectively. (Note: the expected disposition tables do not include the nonresponse follow-up sample.)

Table 1: Expected Disposition for PAC User RDD Survey

Total Purchased Sample		71,715		
held in reserve	20%	14,343		
main sample (released)	80%	57,372		

Determining HH Status			57,372	
	HH status not ascertained	**10%**	5,737	
60%	not a HH	54%	30,981	
40%	HH to screen	36%	20,654	

Screen for Eligibility			20,654	
	Unknown	**65%**	13,425	
90%	Not eligible	32%	6,506	
10%	Eligible	4%	723	

Interview Status			723
Not interviewed	17%	123	
Completed interview	**83%**	**600**	
		Overall Response Rate	*26%*

Table 2: Expected Disposition for PAC User Cell Phone Survey

Total Purchased Sample		9,450		
held in reserve	20%	1,890		
main sample (released)	80%	7,560		

Determining HH Status			7,560	
	HH status not ascertained	**15%**	1,134	
50%	not a HH	43%	3,213	
50%	HH to screen	43%	3,213	

Screen for Eligibility			3,213	
	Unknown	**70%**	2,249	
90%	Not eligible	27%	867	
10%	Eligible	3%	96	

Interview Status		96
Not interviewed	17%	16
Completed interview	**83%**	*80*
Overall Response Rate		*21%*

Both the RDD and cell phone samples will be randomly partitioned into sample replicates and released/managed separately. This protects against unexpected values in our key design parameters (e.g., the eligibility rate is 40% higher than expected). The data collection will commence slowly using only a few replicates. Based on the performance of these early replicate releases sample and design parameters will be fine tuned in order to reach the expected targets without exhausting the budget. Because of the large number of households that need to be screened and the limited availability of funds, advance letters will *not* be issued. To mitigate this, a minimum of 10 call backs to each sampled household will be used with calls staggered across different times and days of the week. Also, replicates will be given a 'rest' (i.e., not called for several days to a week) after 6 attempts after which time they will be re-fielded.

We expect the screening questions will average 2 minutes for administration; for the eligible respondents, the survey will average 15 minutes to administer. We expect the overall field period to range 10 to 12 weeks.

In accordance with section 3.2 of the OMB guidelines, all response rates will be calculated using weighted and unweighted measures, and item response rates will also be calculated to account for item non-response. For the RDD survey, we anticipate achieving an overall response rate of 26% (see table 1); for the cell phone survey component we anticipate 21%. In the web survey, we expect approximately 80% of selected public library administrative units to participate by linking to the survey on their library homepages. We expect approximately 12.5% of public access computer users to complete the survey. Thus, the expected overall response rate is 10%. Since we are projecting a response rate well below 80% for both samples, we will conduct a non-response study to determine how non-respondents differ from respondents. This is discussed below.

Nonresponse follow-up. The nonresponse follow-up study will be used to explore nonresponse bias stemming from the low response rates in the surveys. The subsamples for the nonresponse follow-ups will exploit the replicated sampling feature of the sample design. The subsamples will consist of the initial replicate release from both the RDD and Cell Phone samples (comprising about 20% of both overall samples). For the non-responders from these initial replicates, mailing addresses will be obtained to the extent possible using commercially available reverse matching services (e.g., Telematch, Equifax). For those telephone numbers with addresses, we will send a nonresponse notification letter offering the option of completing survey online or by phone, and offering an incentive for completion. After a 'resting period' we will call and repeat an incentive offer for participation, including an incentive for screening. Our goal is to reach an overall 40% response rate for this nonresponse follow-up (regardless of frame), yielding an incremental 80 interviews from the follow-up replicates (i.e., beyond the number that will have been produced under our 'usual protocols' prior to the launching of 'follow-up activity').

In accordance with section 4.1 of the OMB guidelines, and in order to reduce non-response bias and increase the value of survey data, the final sample will be post-stratified to match national parameters for sex, age, education, race, and Hispanic origin, as taken from the U.S. Bureau of the Census. We will employ CPS or ACS population controls (whichever are most timely and appropriate).

After taking into consideration the weighting and survey design effects, we expect that the sample size for the telephone survey (760 users) will result in a margin of error of about +/-4.5% with a confidence level of 95%. We use a design effect of 1.6 to reflect the impact of differential weighting.

B2.2 Web survey

Within each sampled library system, selected libraries will be contacted through their state library agencies. The Chief Officers of State Library Agencies (COSLA) has agreed to help coordinate the web survey effort by contacting public libraries. We expect that approximately 80% of selected libraries will participate by hosting a link to the web survey on their library's web site during a two week data collection period. All public access computer users in those libraries will be eligible to respond to the survey.

The sample of 400 library systems will be divided into replicates and tracked separately. The data collection will start slowly using only a few replicates. Based on the performance of these early replicate releases sample and design parameters will be fine tuned in order to reach the expected targets without exhausting the budget We anticipate replicates every two weeks over a two month period. This methodology also helps to account for seasonal variation in PAC usage.

The main purpose of the web survey is to generalize results at the library and community level—that is, to understand how characteristics of the library and the community it serves affects the individual public access computer user's experience. It will also provide us with the opportunity to gather data from low income or homeless persons who may not own telephones and respondents in the 14-17 age range who may be disproportionally missed in the telephone survey.
Although individual survey participants will largely be self-selected, i.e. they will follow a link to the survey when they engage in a public access computer session, we will be able to estimate the non-response rates by tracking the number of survey responses against the library's reported use of their public access computer terminals, [2] as well as quit-rates as recorded through the survey web site. While non-response rates to web surveys have been demonstrated to be significantly higher than other survey modes, this may not hold true for the PAC user population. Unlike other web survey environments, PAC users will be accessing the web free of charge at the time they are presented with the opportunity to participate in the survey. The 'free good' of the PAC serve may trigger a reciprocity reaction to participate; moreover, the salience of the survey may be higher among this specific subpopulation than for the general public.[3]

---

[2] Libraries ordinarily collect data on public access computer use.
[3] Groves, R.M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly, 70(5), p.646-675.*

Nonresponse Studies. There are two levels of nonresponse in the Web survey – (1) nonresponse at the PSU level, from library systems declining or otherwise unable to participate; and (2) nonresponse from sampled PAC users who decline to participate in the self-administered web survey when they present at a PAC station where data collection is occurring. We will address the PSU (library system) level nonresponse through an analysis of our rich sampling frame data which contains a variety of PAC and other characteristics for each library system in the U.S. As needed we will use the frame data to identify characteristics that correlate to both response propensity and items related to PAC service/usage. These will then be used to develop post-stratification adjustments at the library system level.

A nonresponse study is also planned for the second level of sampling – PAC users. Since it is not feasible to follow-up those individuals who decline to participate when they commence a PAC user session, our strategy is to use a subsample of libraries to perform a more rigorous data collection operation that would not be practicable for the group of 400 participating library systems. Our approach will be to select a (stratified) random sample of 15% of the participating library systems (with about 400 expected participating library systems, about 60 will be subsampled). These library systems will be the focus of a more rigorous, costly field protocol that is expected to produce a significantly higher response rate than the balance of the PSUs. In these library systems, a $200 stipend will be provided to support the training of librarians as field representatives on behalf of the survey. Their role will be to actively encourage PAC users to participate through one-on-one contact. We anticipate this will at least double the response rate at the user level to 25 percent (from 12.5%). The substantial differential response rate will allow the exploration of potential nonresponse bias, as well as identify correlates of both response propensity and user characteristics. This will allow for a more informed, judicious selection of nonresponse adjustments (e.g., we may find that individuals' response propensities and demographic characteristics vary by time of day and day of week, suggesting that nonresposne weighting adjustments using these factors would reduce potential bias. We note that extant data from other surveys (e.g., CPS, ACS) are not available for post-stratification, and this is a recognized limitation of the data.

B2.3 Case studies

Selected case study libraries will be sent a letter informing them of the study and requesting their participation. Libraries will then be contacted by telephone to arrange the local site visit. The initial telephone contact will provide background about the project and seek additional information on organizations and partners in order to identify key respondents. Based on this information, we will contact respondents and determine the best timing for the visit in order to accommodate the schedule of local respondents.

The case study site visits will be conducted by two-person teams drawn from University of Washington graduate students and researchers. Each team will be composed of one senior and one junior researcher. Senior staff on this project are experienced in field-based qualitative research and semi-structured interviewing of the type that will be used in this study. All researchers involved in the fieldwork will be trained with respect to the objectives of the study and the procedures to follow during the site visits.

B3. Methods to maximize response rates and to deal with issues of non-response

B3.1 Telephone survey

Methods to maximize the response rate for the national sample commence with the pretesting activity that precedes the launch of data collection. During the debriefing of respondents, we will explore factors related to the decision to participate so that we can enhance the field protocols at both the screening introduction as well as the introduction to the substantive questions. Moreover, we will conduct debriefings with the telephone interviewers after the pretest to gather feedback on how to best gain cooperation and avert break-offs. We will launch the survey using highly trained interviewers that will undergo an especially rigorous training and careful monitoring, all of which incorporate the findings of the pretest. Up to 10 attempts will be made to complete an interview for every sampled telephone number. These calls will be staggered over times of day and days of the week and rested to maximize the chances of making contact. Finally, respondents will be offered a $10 Amazon.com gift card. Although resources do not allow for the issuance of advance letters and advance incentives (nor are these appropriate given the low eligibility rate of households), we expect that our proposed approach will maximize the overall response rate to the telephone survey given the essential survey conditions that must be adopted.

B3.2 Web survey

Like the telephone survey, methods to maximize the response rate for the web survey will commence with the pretesting activity that precedes the launch of data collection. Our early site visits to library systems will explore how to best approach and secure participation from the sampled library systems. Specific attention to 'usability' and comprehension will be important foci of the web testing. Our testing of PAC users will include a debriefing component that also gathers feedback on optimal approaches to securing participation, including visual cues (size, color placement, design) and navigation burden on the web screens, utility of written material on the survey (e.g., study brochure), and availability of library staff to answer questions.

In terms of field preparation, State library agencies will provide assistance deploying the web survey to the randomly selected public libraries in their states. Links to the web survey will be prominently displayed on public library websites. Public access computer users will see the link on the library's home page when they open an Internet browser. Participating libraries will be provided with appropriate tools and assistance to link to the survey, including buttons, float-in windows, and news items, as well as detailed instructions on how to use them. A wiki site will be maintained to help selected libraries communicate with the research team and resolve any problems. Participating state agencies will receive $2,000 in order to maximize their participation and participating library systems will receive a customized analysis and report of the data collected at their site.

The questions on the web survey will be designed so that respondents will not need to reference outside resources that would take additional time, and skip patterns will be used through the web-based technology to help move the respondent through the questions in an efficient manner. The online survey will eliminate the need for hand entering paper survey responses and will

achieve the best balance between maximizing data quality and controlling measurement error while minimizing respondent burden and cost.

B3.3 Case studies

State librarians will aid in the selection of public library sites where we will conduct case studies. Public libraries participating in case studies will be given $5,000. Library users who participate in interviews or focus groups will be given a $20 incentive for their participation.

B4. Tests of procedures or methods

The research procedures have been reviewed by the UW Internal Review Board and comply with federal regulations regarding the protection of human subjects participating in academic research. Subjects will be informed of their rights as participants, and may refuse to answer any question or end their participation at any time. All responses will be anonymous; no identifying information will be recorded. Subjects will be at minimal or no risk of suffering stress, embarrassment or discomfort from this study. Children under age of 14 are not competent to give legal assent, thus ineligible to participate.

B4.1 Telephone survey

Prior to launching, the instrument will be pre-tested with 100 people in four different Seattle-area communities. Individual survey items will be reviewed for clarity, and item responses will be analyzed for internal reliability (i.e. Cronbach's alpha). As indicated earlier, the pretest will also be used to gather information to enhance participation at both the screening and interview steps of data collection. As part of the pretest we will also review the captured data to ensure that the applications are functioning as intended and to identify potential problem questions (e.g., those with very little or no variation in response distributions).

Once survey data collection is launched, we will review the survey data to check for data-entry errors, inconsistencies, and identify missing cases for any systematic bias. Cases with missing responses on half or more of our pre-specified key survey items will be deleted from the sample and treated as a unit nonresponse. Initial analysis will consist of running descriptive statistics for all variables to identify the center and distribution within the population and bivariate statistics (correlation and cross-tabulation) will be used to test for associations between variables. For variables where the team identifies an association based on the qualitative evidence, we will conduct multivariate analysis (e.g., multiple regression) to determine the proportion of variance that can be explained by the relationship.

B4.2 Web survey

We will conduct a small pilot study that tests the web survey prior to the main data collection effort. Pre-testing will occur in 5 Seattle-area branch libraries. This number is likely to generate a pre-test sample of 100-200 responses, which is adequate for internal reliability testing and cross-validation to telephone survey results. The data will also be checked to establish the integrity of

the computer application to capture the recorded responses as intended. The web survey test will also focus on issues of usability and survey participation (as discussed earlier).

After collection, we will conduct data assessments and other reliability testing as described for the telephone survey. In addition, we will use comparisons between the data sets to detect unaccounted for non-response error in the web survey group.

B4.3 Case studies

A initial pre-test of 9 participants of each respondent type was conducted at a Washington library to test interview guides and field procedures. Results from this test will be used to refine field protocols, instruments, and training for investigators, and to begin developing the codebook. In general, library administrators, employees, and patrons were all found to be extremely eager to share their experiences with public access computers.

We will analyze case study data as they are collected in order to aid in identifying a range of responses for each indicator. The schemes will reflect the data's emergent themes and will be guided by the study's logic model. The code book will be used to assign terms to all segments in the data that reflect particular concepts. After the final schemes are developed, tests of inter-coder reliability (c.f., Krippendorf, 1980) will be conducted with independent coders and final adjustments will be made to the codes. (Inter-coder reliability testing will be conducted during analysis of the open-ended survey responses prior to collection and analysis of case study data.)

To ensure trustworthiness (reliability and validity) of the qualitative data, we will use several measures (c.f., Chatman, 1992; Lincoln & Guba, 1985). Reliability will be ensured through: (1) consistent note-taking, (2) exposure to multiple and different situations using triangulated methods, (3) comparing emerging themes with findings from related studies, (4) employing intracoder and intercoder checks, and (5) analyzing the data for incidents of observer effect. Validity will be assessed as follows:

- Face validity: ask whether observations fit an expected or plausible frame of reference;
- Criterion/internal validity (credibility) based on pre-testing instruments, rigorous note-taking, methods, peer debriefing, and member checks or participant verification;
- External validity: provide "thick description" and comprehensive description of our methods so others can determine if our findings can be compared with theirs;
- Construct validity: examine data with respect to public access computing outcome literature, models of public library use, and principles of information behavior

B5. Individuals consulted on statistics and on collecting and/or analyzing data

The agency responsible for funding the study, determining its overall design and approach, and receiving and approving contract deliverables is:

U.S. Institute of Museums and Library Services
Office of Policy, Planning, Research and Communications

1800 M Street NW, 9th Floor
Washington, DC 20036-5802

Phone: 202-653-4630

Person Responsible: Mamie Bittner

The University of Washington, Information School is the prime cooperator for this study. It
is responsible for implementing the overall design of the study and development of the data
collection instruments. It will also field the case studies using its own staff, and will have
responsibility for all data analyses obtained through the telephone survey, web survey case
studies, and focus groups.

The Information School
Box 352840
Mary Gates Hall, Ste 370
Seattle, WA 98195-2840

Phone: (206) 685-9937
Fax: (206) 616-3152

Persons Responsible: Karen Fisher and Mike Crandall, principal investigators and Matthew
Saxton, survey methodologist and statistical expert

**The Urban Institute was consulted in the development of the telephone and web survey
sampling frames and follow-up study methodology.**

2100 M Street, N.W.
Washington, D.C. 20037

Phone: 202-833-7200

Persons responsible, Robert Santos, Senior Institute Methodologist and Timothy Triplett Survey
Methodologist

Telephone Contact, Inc. will conduct the telephone survey.

Telephone Contact, Inc.
3800 Hampton Ave., Ste. 200
St. Louis, MO 63109

Phone: 314-353-6666

Person responsible: Joyce Aboussie, President and CEO

The web survey will be administered by the Survey Research Division of the University of Washington. It will be responsible for the programming of the web survey device, the storage and security of data, and technical advice on linking mechanisms.

Survey Research Division
Social Development Research Group
University of Washington
9725 3rd Ave NE, Ste 401
Seattle, WA 98115

Phone: 206-616-9642

Person responsible: Kimberly Cooperrider, Technology Director

References

Chatman, E. A. (1992). *The information world of retired women*. New York: Greenwood Press.

Chute, A. and Kroe, P.E. (2007). Public Libraries in the United States: Fiscal Year 2005 (NCES 2008-301). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Newbury Park, CA: Sage

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.

Manfreda, et al. (2008). "Web surveys versus other survey modes: A meta-analysis comparing response rates." *International Journal of Market Research*, 50(1).