

U.S. Department of Education

Post Vocational Rehabilitation Experiences Study

*Office of Management and Budget
Clearance Package Supporting Statement
And Data Collection Instruments*

Part B

May 16, 2007

Revised August 8, 2007

Contents

	<u>Page</u>
Introduction.....	1
A. Justification.....	2
A1. Circumstances Making the Collection of Information Necessary.....	2
A2. Purpose and Use of the Collected Information.....	7
A3. Use of Technology to Reduce Burden.....	9
A4. Efforts to Identify Duplication.....	10
A5. Methods to Minimize Burden on Small Entities.....	11
A6. Consequences if Collection is not Conducted.....	11
A7. Special Circumstances.....	11
A8. Federal Register Comments and Consultations Outside the Agency.....	12
A9. Decision to Provide Any Payment or Gift to Respondents, Other Than Remuneration of Contractors or Grantees.....	17
A10. Assurance of Confidentiality.....	17
A11. Justification for Any Questions of a Sensitive Nature.....	20
A12. Estimates of the Hour Burden.....	21
A13. Estimate for the Total Annual Cost Burden to Respondents.....	23
A14. Estimates of Annualized Costs to the Federal Government.....	23
A15. Reasons for Any Program Changes or Adjustments.....	24
A16. Plans for Tabulation and Publication, Analytic Techniques, and Time Schedule.....	24
A17. Approval to Not Display the Expiration Date.....	28
A18. Explanation of Exceptions.....	28
Part A References	29
B. Collections of Information Employing Statistical Methods.....	32
B1. Respondent Universe and Sampling Methods.....	32
B2. Procedures for the Collection of Information.....	36
B3. Methods to Maximize Response Rates.....	41
B4. Tests of Procedures or Methods.....	53
B5. Consultations on Statistical Aspects of the Design.....	54
Part B References.....	55

List of Tables

	<u>Page</u>
Table 1	Estimates of VR Agency Respondent Burden – Baseline Only.....21
Table 2	Estimates of Survey Respondent Burden.....22
Table 3	Estimated Annualized Costs to the Federal Government.....23
Table 4	Estimated Costs by Expense Category and Study Phase24
Table 5	Types of Data Analyses and Usefulness27
Table 6	Response Rates for Ticket-to-Work Study and the SSI/Medicaid Surveys and Projected Response Rate for PVRES35
Table 7	Definition of Sampling Strata and Sample Allocation.....37
Table 8	Allocated Sample Sizes and Expected Standard Errors in the Second Followup by Subgroup38

List of Exhibits

Exhibit A-1	Panel of Experts.....15
Exhibit A-2	Anticipated PRVES Schedule.....28

List of Appendixes

	<u>Page</u>
Appendix A. Crosswalk between Study Questions, Data Elements and Interview Questions.....	A-1
Appendix B. Survey Instrument.....	B-1
Appendix C. Federal Register Notices.....	C-1
Appendix D. Advance Materials.....	D-1
Appendix E. Confidentiality and Non-Disclosure Forms	E-1
Appendix F. Request to State Agencies	F-1
Appendix G. List of Non-Response Bias Study Variables.....	G-1
Appendix H. Refusal Confirmation Script.....	H-1

**SUPPORTING STATEMENT FOR
PAPERWORK REDUCTION ACT 1995 SUBMISSION
INFORMATION COLLECTION PLAN FOR THE
POST VOCATIONAL REHABILITATION EXPERIENCES STUDY**

B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS

B1. Respondent Universe and Sampling Methods

Potential Respondent Universe

PVRES is a survey of former VR consumers, whose case records were closed after receiving services under the State Vocational Rehabilitation program, including those who achieved an employment outcome at case closure and those who did not. The study focuses on the four subgroups of the VR consumers described in A2: MR, MI, TY, and SSB. It is a longitudinal study, which calls for three waves of data collection. The total sample size is set at 8,000 in the first wave.

The four subgroups of the study constitute their own universes rather than domains of a bigger study population. However, studying each subgroup independently would be more costly than studying them together in one survey. This approach defines a universe that includes all former VR consumers who received VR services and had at least one of the subgroup characteristics.

Each fiscal year, states, the District of Columbia, and territories report to RSA all VR cases closed in the fiscal year. RSA compiles the reports and produces the RSA 911 for each fiscal year. The RSA 911 file for Fiscal Year (FY) 2006 will provide the most recent sampling frame for the PVRES study to be conducted in 2007. The study universe is defined by the data file to be used as the sampling frame. Therefore, there is no coverage issue due to an imperfect frame.

A restricted version of the FY 2006 RSA 911 file that contains no personal identifiers was used to develop the sampling plan. There were 345,899 former VR consumers whose records were closed after receiving services, with or without employment, in the FY 2006 RSA 911 file. Among them were 213,039 former VR consumers who had at least one of the four

subgroup characteristics. So, the PVRES universe represents about 62 percent of the whole of former VR consumers who received services and exited with or without employment.

The sampling frame will include the case records for members of the four subgroups of interest that were closed in FY 2006 after receiving VR services. To identify only individuals who received services and exited with or without an employment outcome at closure, Item 36 in the RSA 911 file, Type of Closure, will be used, where Type of Closure = 3 (exited with an employment outcome) or 4 (exited without an employment outcome). Excluded will be individuals served by VR agencies located in territories, those whose Reason for Closure (Item 37) is 04 (death), and former consumers who do not belong to the four subgroups of interest. There will be some individuals whose services were closed more than once in the same fiscal year. In these cases of multiple records for the same individuals, only the latest closures are kept.

In the RSA 911 file, the subgroup members are identified by Item 13, the Primary Disability of a VR consumer. The primary disability is defined as the primary impairment that causes or results in a substantial impediment to employment. The RSA 911 file also includes the secondary disability, the physical or mental impairment that contributes to, but is not the primary impediment to employment. Secondary disability will not be used in the formulation of the sampling plan.

The MR and MI subgroups are defined using Item 13 (Primary Disability), which is a 4-digit code composed of two parts. The first 2 digits indicate the impairment type, and the third and fourth digits indicate the impairment cause. Mental retardation is indicated by an impairment cause code of 25 (Mental Retardation). Mental illness is defined similarly by impairment cause code, including: 04 (Anxiety Disorders), 15 (Depressive or other Mood Disorders), 24 (Mental Illness not listed elsewhere), 29 (Personality Disorders), and 33 (Schizophrenia and other Psychotic Disorders).

The TY subgroup is defined as VR consumers with an age greater 13 but less than 22 at time of application and an age less than 33 at time of closure. The age at application is determined using Items 5 (Date of Birth) and Item 6 (Date of Application) and the age at closure is determined by Item 5 and Item 38 (Date of Closure).

The SSB subgroup is defined using Item 18 (Type of Public Support at application). Receipt of Social Security benefits is defined as Item 18 having a value of Supplementary Security Income or a value of Social Security Disability Insurance.

Expected Response Rates for the Collection as a Whole

The response rate projection is obtained following recommendations of members of the POE having experience surveying persons with disabilities. Among these recommendations was to use the experience of two recent studies of persons with disabilities as the basis for PVRES response rate projections. These are the Evaluation of the Ticket-to-Work Program (Thornton et al., 2006) and the SSI/Medicaid surveys conducted for the Evaluation of Section 1115 Medicaid Reform Demonstrations (Mitchell et al., 2006). As evidenced by these and other studies, the PVRES study population, which includes a large percentage of individuals with mental retardation or other mental illness, is a very challenging group to reach and successfully complete an interview. The experience of these two studies is provided in the second and third columns of Table 6.

The Ticket-to-Work study was based on a national sample of disability beneficiaries drawn from SSA's administrative records. The study used two interview modes, computer-assisted telephone interviewing (CATI) augmented by computer-assisted personal interviewing (CAPI) for CATI nonrespondents. The SSI/Medicaid samples included people with physical and sensory disabilities, mental illness, and mental retardation in three states. It used CATI data collection methods exclusively.

Since the Ticket-to-Work study was a national study but the SSI/Medicaid surveys only covered a few states, we believe that PVRES is more like the Ticket-to-Work study than the SSI/Medicaid surveys. Therefore, the response rate for the CATI mode of the Ticket-to-Work study (63 percent) is assumed to be the response rate for PVRES at the baseline. The relative response rates of the subgroups are modeled after the SSI/Medicaid surveys because such a breakdown is not available for the Ticket-to-Work study. Based on the population distribution of the three strata the response rate difference between Physical/Sensory and the two mental disability groups in the SSI/Medicaid surveys, the response rates should be 65 percent (non MR/MI) and 60 percent (MR/MI) to arrive at the overall response rate of 63 percent. Applying these rates to the PVRES study sample distribution, the overall response rate of 61 percent at the baseline is obtained (Table 6).¹

¹ If the sample drawn from the FY 2006 RSA 911 is composed of 74 percent of consumers with MR or MI, the overall response rate will be 61.3 percent ($= 60(74/100) + 65(26/100)$).

Table 6. Response Rates for the CATI Part of the Ticket-to-Work Study and the SSI/Medicaid Surveys and Projected Response Rate for PVRES

Disability Type	Ticket-to-Work CATI ²	SSI/Medicaid	PVRES	
			Baseline	2nd Followup
Physical/Sensory ¹	NA	69.6	65	52
Mental Illness	NA	64.5	60	48
MR/DD	NA	64.0	60	48
Unknown	NA	65.3	NA	NA
Total	63	65.7	61	49

¹ Note: For PVRES, this category contains all other non-MI and non-MR consumers.

² Note: The study had 6,302 CATI completes out of 9,999 eligible sample units (see Thornton et al., 2006, pp. C-2 and C-5).

While we anticipate lower than average response rates for the baseline, the expectation is that once consumers are located and their participation secured, a good response rate can be obtained at each of the two followup surveys. The response rates for the first and second followup surveys are projected to be 85 and 80 percent of the baseline respondents, respectively.

The response rate for PVRES faces the additional challenge that cases were closed for about 13 percent of the former consumers in the FY 2006 RSA 911 data file because the consumer could not be located by the VR agency (Reason for Closure equals 01 on Item 37). Because little is known about why these persons could not be located, PVRES will include these consumers in the sample frame and attempt to locate as many as project resources will allow. The contact information that will be obtained through the state VR agencies for these cases will be the same information the agencies had when the cases were closed as not located. Lack of viable contact information for this subset of cases will have a negative affect on the overall response rate.

The following steps are being taken because a response rate lower than 80 percent is projected for the baseline data collection:

- Extended steps will be taken to maximize the response rate. These are described in detail in B3 under “Methods to Maximize Response Rates.
- A nonresponse analysis will be conducted. This is described in the response to question B3 under the heading of “Dealing with Nonresponse.”
- The adequacy of the estimates for purposes of the study is examined and justified in B2: Procedures for the Collection of Information.

To reach a response rate of at least 80 percent would require more extensive and costly activities than has been allocated for this study. Cases that are not completed by telephone or paper (about 39% of the sample) would have to be referred to in-person tracers and interviewers. This was done in the National Beneficiary/Ticket-to-Work survey in order to reach people without telephones or individuals who could not use a telephone. For PVRES, field interviewers would search neighborhood, transient housing and shelters. They would also speak to group home administrators and others who might be more willing to provide information in person than over the telephone. These informers could help identify guardians and proxy respondents. This type of tracing is very labor intensive and would extend the data collection period to reach the non-respondents.

For PVRES, about 3,000 cases would be eligible for referral to in-person tracing and interviewing. To reach an 80 percent response rate would require locating and interviewing about 1,500 of the 3,000 cases. Because PVRES is not a clustered sample (appropriate information for clustering is not in the RSA 911 sampling frame), the field staff would need to be large, their assignments geographically dispersed, and the field period extended significantly. The extension would introduce disparities into the reference period of the interviews, affecting the analysis.

B2. Procedures for the Collection of Information

The desire is to achieve the maximum precision level for estimating the employment rate for each subgroup with the given initial sample size of 8,000. This calls for stratification of the study universe by subgroup characteristics. MR and MI are the only mutually exclusive characteristics. TY and SSB will overlap with each other as well as with MR or MI. As a result, there are 11 strata defined by the various combinations of these subgroup characteristics. These are shown in Table 7 with population distribution.

Allocation of the total sample of 8,000 to the 11 sampling strata is performed in order to achieve the maximum level of precision for an estimate of the employment rate of each of the four subgroups at the last wave of data collection (the second followup). Because the employment rates at case closure ranged between 40 and 60 percent across the four subgroups, we assume 50 percent for all subgroups for the calculation of the precision level. This assumption provides a conservative sample allocation in the sense that the allocated sample is always large enough to satisfy the precision requirement with some margin, provided the projected response rates are achieved.

Table 7. Definition of Sampling Strata and Sample Allocation

Stratum Type	MI = 1 MR = 2 Not MR or MI = 0	TY = 1 Not TY = 0	SSB = 1 Not SSB = 0	Stratum Number	Population size	Sample size
SSB only	0	0	1	1	43,312	844
TY only	0	1	0	2	52,523	1,269
TY, SSB	0	1	1	3	9,330	218
MI only	1	0	0	4	35,689	1,351
MI, SSB	1	0	1	5	25,637	940
MI, TY	1	1	0	6	9,648	438
MI, TY, SSB	1	1	1	7	2,341	103
MR only	2	0	0	8	5,582	423
MR, SSB	2	0	1	9	11,124	814
MR, TY	2	1	0	10	11,146	1,011
MR, TY, SSB	2	1	1	11	6,707	589
Total					213,039	8,000

Although the response rate for the MI and MR groups is lower, the same level of precision for all subgroups could be obtained because the population distribution over the sampling strata is more favorable to MI and MR groups. One reason is that MI and MR subgroups each cross with only two other subgroups (TY and SSB) whereas the TY and SSB subgroups each cross with three other subgroups (e.g., MI, MR, and SSB for TY). If a subgroup sample scatters over many strata with differential sampling rates, the sample becomes less efficient because the sampling weights vary more; this results in a higher design effect. The resulting allocation based on FY 2006 data is shown in Table 7.

Simple random sampling will be used to select a sample of former VR consumers from each stratum, and so the PVRES sample design will be a stratified simple random sample. Table 8 shows the expected standard error for an estimate of the employment rate in the second followup for each subgroup under the assumption of projected response rates given in Table 6 and employment rates of 50 percent. This level of precision is considered adequate for the purpose of the study.

Table 8. Allocated Sample Sizes and Expected Standard Errors in the Second Followup by Subgroup

Subgroup	Population Size	Baseline Sample Size	Second Followup Sample Size	Second Followup Standard Error
MI	73,315	2,831	1,359	1.35
MR	34,559	2,837	1,362	1.35
TY	91,695	3,628	1,887	1.35
SSB	98,451	3,509	1,825	1.35

The table also provides the allocated initial sample sizes by subgroups, which are obtained from Table 7. When the sample includes TY and/or SSB consumers, there is some overlap of TY or SSB status with the MI or MR status. Therefore, the baseline sample sizes do not sum to the total sample size of 8,000 in Table 8. In contrast, the MR and MI groups do not overlap.

Estimation Procedure

The base weight will be calculated for the baseline sample as the reciprocal of the inclusion probability of each sample unit. The base weights will be adjusted for unit nonresponse creating weighting cells with help of the Chi-squared Automatic Interaction Detection (CHAID) analysis using primarily design variables, and variables of age category, gender, and disability category. The CHAID analysis is discussed more fully in the response to B3.

The nonresponse adjusted weights will be further modified to ensure the weighted sum of each subgroup sample equal to the known subgroup population size.

To facilitate variance estimation, the jackknife variance estimation method will be used. Random groups will be created within design strata to form clusters for variance estimation; this will help to create a manageable number of replicates.

Since the study is longitudinal, different weights will be developed in each wave of the baseline and followup surveys. At the baseline, only one set of weights will be produced, but at the end of each followup data collection period, two different sets of weights are needed and will be developed: one for cross-sectional analysis and the other for longitudinal analysis. For developing the cross-sectional weights, we will use all respondents for the followup wave. For longitudinal analysis the longitudinal weights will be developed by adjusting the baseline

weights for wave nonresponse using the same adjustment cells used for cross-sectional weighting as much as possible.

Software such as HLM6.0, which is used for hierarchical linear modeling, can accept cases with missing waves as long as they have values for at least two waves. If it is deemed necessary to use this feature, another set of longitudinal weights will be developed in the last followup for the longitudinal sample of cases with at least two waves' data.

All analyses will be performed using the appropriate survey weights developed as described above.

Unusual Problems Requiring Specialized Sampling Procedures

There are no unusual problems requiring specialized sampling procedures.

Questionnaire Design

Sampled former consumers will be interviewed three times: baseline survey and two annual followup surveys. The CATI and paper versions of the data collection instrument appear at Appendix B. Questions in the interview have been borrowed from or adapted from related research where possible.

The data collection instrument contains questions to be asked at baseline and at followup. Demographic information that is not subject to change will be obtained only once, either from the RSA files or at the first interview. Other information covered by the questionnaire is subject to change from interview cycle to interview cycle, and we propose to update it at each followup. The baseline interview will ask the former consumer about experiences since leaving vocational rehabilitation or during the last 12 months, whichever is more recent. Each followup interview will ask about the time period that has elapsed since the last interview (typically 12 months).

Data Collection Methods and Procedures

The data collection procedures will be adapted to the needs of a disabled population to enhance the response rate and address issues of data quality. These procedures must provide the ability to communicate with all members of the study population and they must be sufficiently flexible to work around issues of both physical and mental fatigue. The interview must be able to communicate effectively with the cognitively impaired.

The PVRES interview will have the following attributes to address issues of accessibility and quality of response.

- The interview will be available under two modes: telephone and paper. Respondents will be encouraged to participate via the mode with which they are most comfortable. TTY and relay services including voice-carry-over, hearing-carry-over or speech-to-speech and respondent-provided amplification will be offered to the hearing impaired. Respondents will be encouraged to utilize accommodations they typically use to access information via telephone or mail.
- Telephone interviewers will be given alternative wording that may be used to simplify and more clearly state complex questions for the cognitively impaired. They will be given alternative wording that relies on different sounds for the hearing impaired. (These appear as PROBES in the telephone version of the instrument.)
- The telephone interview will be capable of being administered in multiple sessions to accommodate persons with physical impairments who may become fatigued or those with mental impairments that limit the ability to concentrate.
- Interviewers will follow criteria to determine when an interview should be conducted by proxy and will have procedures for identifying an appropriate proxy. Records will document when the interview was conducted by proxy.
- The paper survey has been designed to minimize the need for respondents to follow complex “skips” in order to be cognitively more accessible. This resulted in some abbreviation in the interview content in the paper version due to omission of a limited number of detailed followup questions that would not be applicable to all respondents.

Interviewers will receive special training to prepare them for issues that will arise when interviewing persons with disabilities. In addition to general telephone interviewer training, nonresponse avoidance and conversion training, and training in the specific content of the interview, PVRES interviewers will receive sensitivity training. This training will focus on preparing staff for what to expect and techniques to use in different situations. Interviewers will also be prepared to deal with a variety of special circumstances that may arise, such as:

- Working through guardians or other gatekeepers;
- Conducting interviews of persons in institutional settings; and
- Identifying the need for and arranging for proxies, translators, or facilitators.

Interviewers will begin calling sampled consumers within 10 days of the advance packages being mailed. When they contact a prospective respondent, they will evaluate the need for a proxy by administering a cognitive test, administer the informed consent introduction, and determine requirements for accommodations. The contact and screening scripts are provided in Appendix F. Various tracing methods (described under B3) will be used to locate former consumers no longer at the address or telephone number contained in the state VR records.

Persons who call in to refuse to participate or who refuse by completing and returning a signed Informed Consent Form will be contacted by telephone to confirm their decision. (See Appendix H for confirmation script.) Persons who request a mail survey and do not return it within three weeks will be contacted by telephone and encouraged to respond (by the mode of their choice).

Any Use of Periodic (Less Frequent Than Annual) Data Collection Cycles to Reduce Burden

There is no use of periodic data collection cycles to reduce burden.

B3. Methods to Maximize Response Rates

Overall response projections were presented earlier. Achieving this response rate involves locating the sample members and securing participation. Those completing a baseline interview will be eligible for the first and second followup. We estimate 85 and 80 percent of the respondent consumers at the baseline will complete the first and second follow-ups, respectively, yielding a 49 percent final response rate. (Earlier discussion of the sample precision indicated the adequacy of this response rate for the intended analysis and reporting.)

There are two key aspects to maximizing the number of sample members for whom data are collected: completing data collection with the maximum number of sample members who are retained in the sample and minimizing the number of sample members lost through attrition.

We discuss minimizing attrition later. Here we describe procedures to be followed to maximize the number of sample members who complete the survey:

- Former consumers will have the option of completing the survey using the mode of their choice (telephone or mail).

- Consumers with known email addresses will receive reminders via email to complete the survey.
- We will follow up by telephone with all consumers who do not complete the survey within a specified period.

The following procedures will be used to maximize the completion rates for surveys that are administered by telephone.

- Use a core of interviewers with experience working on telephone surveys of households, particularly interviewers who have proven their ability to obtain cooperation from a high proportion of sample members.
- Require all interviewers to successfully complete training specific to this study, including issues that may arise, facilitating response when working with individuals with disabilities, discussions of how to avoid inviting a refusal, approaches that will help in addressing questions respondents are likely to ask, and how to counter objections.
- Allow a greater number of rings per call to afford the disabled more time to answer.
- Use call scheduling procedures that are designed to call numbers at different times of the day and week, to improve the chances of finding a respondent at home.
- Make every reasonable effort to obtain an interview at the initial contact, but allow respondents flexibility in scheduling appointments to be interviewed.
- Train interviewers to identify when proxies should be used and to conduct interviews through proxy.
- Closely supervise interviewers during data collection.
- Conduct silent monitoring of interviews to identify and promptly correct behaviors that could be inviting refusals or otherwise contributing to low cooperation rates.
- Leave a message on answering machines in order to let the respondent know the call was not a marketing effort but a research study and to accommodate consumers who prefer to call back because they require assistance to use the telephone.
- Send postcards when unanswered calls suggest calls are being screened or when messages on answering machines prove ineffective.

- Provide a toll-free number (voice and TTY) for respondents to call to verify the study’s legitimacy or to ask other questions about the study. Those without telephones in their homes can also call this number from any location and have the interview conducted at that time.
- Require many unsuccessful call attempts to a number without reaching someone before considering whether to treat the case as “unable to contact.”
- Refer “unable to contact” cases to tracing before finalizing the case as “unlocatable.”
- Implement refusal conversion efforts for first-time refusals and use interviewers who are skilled at refusal conversion and will not unduly pressure the respondent.

In addition to encouraging participation in the survey, the offer of incentives has also been demonstrated to be an effective means of reducing survey nonresponse. Sampled consumers will receive a \$10 payment for completing the baseline interview. Respondents to the baseline survey will be retained for followup. The payment will be made regardless of interview mode. It will be made for each respondent in advance for subsequent rounds of data collection. Those who assist respondents or act as proxies will also receive an incentive payment. Interpreters who receive hourly pay will be reimbursed at their regular rate.

Monitor Response

Weekly reports will be produced by the data collection system to assist project staff in monitoring data collection. The Sample Acquisition Plan describes summary reports that will be available regarding enrollment in PVRES. Additionally, summary reports will indicate the status of interviews including

- Number of interim cases by data collection mode;
- Number of finalized cases by final status and data collection mode;
- Number of interim CATI cases by detailed status code;
- Number of interim cases by release wave²; and
- Number of finalized cases by release wave.

² The lead packets used to recruit sampled consumers will be mailed in two waves, referred to as release waves, to balance the workload and ensure all prospective respondents are called promptly after their package is mailed. These waves are described in the Sample Acquisition Plan.

In addition, interviewer statistics reports will be produced that allow the project staff to examine the production, refusal rates, and overall outcome of calls by individual interviewer.

For planning purposes, we have estimated that 9 out of 10 interviews will be conducted on the telephone; the remainder will be obtained via mail.

Debrief Interviewers and Hotline Staff

We will debrief the data collection staff after the baseline interviews and each round of followup. The purpose will be to identify effective data collection techniques that can be shared among data collection staff, determine where training materials might be improved, and identify problems in the survey instrument. Notes of the debriefings will be reviewed by senior project staff who will decide where adjustments are appropriate. Substantive recommendations will be brought to the attention of the Contracting Officer's Representative (COR) and OMB for approval before changes are made.

Conduct Followup Consumer Interviews

Respondents who completed a baseline interview and have not subsequently withdrawn from the study will be eligible to be contacted for each of the two followup interviews. We anticipate that those who enroll in the study during the baseline year will have a high commitment to it. Our expectation is that the response rate for the first and second followup surveys will be 85 and 80 percent, respectively, of those attempted. The combined response by the second followup will be 49 percent of the original sample.

Similar to the baseline data collection, advance letters will be mailed to eligible sampled consumers to begin each of the two followup rounds of data collection. The letters will be mailed to the most current known addresses as determined during panel maintenance activities. Simultaneously, we will send an email message announcing the next round. The email message will be sent to persons who provided an email address during the prior round of data collection. Respondents will have the option of participating by telephone or mail. Efforts will be made to enable respondents to use accommodations that will facilitate their ability to participate in the survey and to provide quality information. This may involve interpreters or proxies.

The computer-assisted interviews will be updated as appropriate to gather followup data. Data needed to drive the interview will be taken from responses to the immediately prior

round where available. In the second followup, we will go back to the baseline interview for information needed to drive the interview for persons who did not respond to the first followup.

Maintain Panel

Longitudinal samples are best maintained if not too much time elapses between contacts and the study stays informed of respondents' current addresses and telephone numbers. PVRES intends to conduct one baseline interview and two followup interviews at one-year intervals. During those intervals between data collection rounds, respondents will be asked for updates to their contact information. These packages will be from RSA and sent first class mail, address-correction requested. Additionally, vendors that specialize in locating services will be used to keep our records current.

No fewer than 3 months prior to each round of followup interviews, we will send letters to all consumers for whom we obtained a completed baseline interview and who did not subsequently ask to withdraw from the study. Prior to sending letters to respondents, we will send the most current information we have on these baseline respondents to the commercial information services that proved productive during sample acquisition. These services will be asked to update addresses and telephone numbers following the same process described for sample acquisition. The Postal Service's National Change of Address database will be part of this effort. We will also match the current sample file against the Death Master File prior to each round of followup data collection. Any updates will be entered in the tracking database along with the source of the information.

The updated tracking database will be used to address letters to the baseline respondents. The letter will remind them of their prior enrollment in the study and of the study's importance. The letter will ask that they contact us by telephone, mail, or email if any of the contact information we have is not current. A toll-free telephone number and the project's email address will be contained in the body of the letter. We will also include a self-addressed postage-paid postcard that can be used to inform us of changes. These packages will be from RSA and sent first class mail, address-correction requested. Any updates we receive will be entered in the study's tracking database and will be available for the next round of data collection.

Dealing with Nonresponse

For item nonresponse, we plan to perform imputation using a software package called AutoImpute. This software has been developed and used internally by Westat for large-scale surveys. The method used by the software can be summarized as follows. For each variable that has missing data and needs imputation, a regression model is first developed using the variable with missing values as the dependent variable and a set of predictor variables that are selected from all available variables in the data set. Predictor variables may have missing values but they are imputed temporarily to be used as predictors. A missing value of the dependent variable is predicted by the model, but the predicted value is not used as an imputed value; rather it is used to form imputation cells by grouping units with similar predicted values. Then a respondent is selected randomly within the imputation cell for a case with a missing value, and its value is donated to the case with missing value. So the imputation procedure is one of donor imputation methods. However, selection of donors is done using an elaborate regression modeling, which pools all available variables in the data set. Another important advantage over other available imputation software is AutoImpute's ability to ensure skip patterns imbedded in the questionnaire.

Handling of unit nonresponse will be done by adjusting the base sample weight. We will use CHAID to create weight adjustment cells for nonresponse. The software examines the relation between the response rate and the size of an adjustment cell to create as many cells as possible while meeting a specified size requirement; the adjustment cells should not be too small so that the resulting adjusted weights are not too volatile. The weight adjustment aims to reduce bias due to nonresponse while controlling variance inflation; nonresponse weight adjustment always increases the variance of a survey estimate by making the weights more variable. We will use design variables (subgroup characteristics that define the sample design strata), and age category, gender, and disability category variables as the starter and proceed to choose a set of the variables that is deemed effective in creation of the weight adjustment cells. A limited number of variables should be used to prevent forming adjustment cells that are too small. We expect that this nonresponse weight adjustment will eliminate most of the nonresponse bias. Because a high nonresponse rate is projected we will verify that this objective has been verified by conducting the extensive nonresponse bias study described below.

The plan for the analysis of nonresponse bias is to use the substantive number of variables available from the sample frame, which are highly correlated with the key survey

variables. These variables will be used to produce population proportions and means from the frame, which will be compared with the corresponding sample proportions and means that will be calculated from the sample respondents and nonresponse adjusted survey weights. It is hoped that the method used to obtain the nonresponse adjustment weights, which will be explained below, will eliminate the bias incurred by nonresponse. If the differences between sample estimates and population values are small, it indicates that the hope may well be realized and nonresponse bias is not a serious threat to valid statistical inferences that will be drawn from the respondent sample with the nonresponse adjusted weights.

The usual approach for handling of unit nonresponse is to adjust the base sample weight of the full sample for nonrespondents. We will use CHAID to create weight adjustment cells for nonresponse. The key to success is to create the adjustment cells that include the sample units with similar response probability. So we need categorical variables that are predictive for the response probability.³ We will use design variables (subgroup characteristics that define the sample design strata), and age category, gender, and disability category variables as the starter and proceed to choose a set of the variables that is deemed effective in creation of the weight adjustment cells.

The CHAID software examines the relation between the adjustment cell response rate and the size of an adjustment cell to create as many cells as possible while meeting a specified size requirement; the adjustment cells should not be too small (the usual cutoff is 20) so that the resulting adjusted weights are not too volatile since volatile weights increases the variance of a survey estimate. The nonresponse weight adjustment using CHAID aims to reduce bias due to nonresponse while controlling variance inflation. For this reason, the number of categorical variables used to form the adjustment cells should be limited to prevent forming adjustment cells that are too small. We hope that this nonresponse weight adjustment will eliminate most of the nonresponse bias. Because a high nonresponse rate is projected we will conduct an extensive nonresponse bias study detailed below.

The RSA 911 data file, which will be used as the sampling frame, provides rich background information for the VR consumers. A list of the RSA 911 variables to be used to study nonresponse is given in Table 1. We believe that some of these variables are highly correlated with key survey variables, so we can use them as proxies to examine whether the sample respondents with the CHAID-based nonresponse adjusted weights would produce

³ If one desires to use a continuous variable, it should be categorized.

unbiased estimates or not. For example, the variable (the 87th variable in Table 1, call it Y) in the RSA 911 file, which provides employment status at closure should be a good predictor for the survey employment status, which is one of the key variables. We can estimate the employment rate at closure using the Y and the nonresponse adjusted survey weights and its standard error. Let the estimate be denoted by p and its standard error by s . Since the whole frame (RSA 911) has the Y -value, we can obtain the true population value denoted by P , which is estimated by p . If the confidence interval constructed using p and s contains the true value P , that is,

$$P \in (p - 1.96s, p + 1.96s),$$

we can say that the estimate p is unbiased for P with a 95 percent confidence.⁴ This will give us confidence in the respondent sample and the nonresponse adjusted weights for estimation of the employment status. Another example is the amount of public supports. The mount at closure (given through the 98th to 101st variables in Appendix G) should be highly correlated with earnings data collected by the survey, and we can test whether or not the respondent sample will produce an unbiased estimate for the mean amount of public supports at closure in the same way. Now the Y -variable is the sum of the 98th, 99th, 100th, and 101st variables in Appendix G. Suppose that a is the sample mean for the Y -variable estimated using the nonresponse adjusted weights and s is its standard error. Further let A be the true mean amount the Y -variable obtained from the RSA 911 frame. If the 95 percent confidence interval includes A as shown below,

$$A \in (a - 1.96s, a + 1.96s),$$

then again we can say with 95 percent confidence that the respondent sample with the nonresponse adjusted weights gives an unbiased estimate for the mean amount of public supports at closure. This will give us some confidence in the sample respondent and the nonresponse adjusted weights for estimating earnings of VR consumers.

We will examine all possible and meaningful variables available in the RSA 911 file in this way. If we can say with a high degree of confidence that the respondent sample with the CHAID-based nonresponse adjustment weights produces unbiased estimates for those RSA 911 variables, we can be confident in the survey estimates produced from the respondent sample with the nonresponse adjusted weights. We will examine more closely the nonresponse bias for those RSA 911 variables, which are considered correlated the key survey variables. If it is judged that

⁴ It is assumed that the distribution of p is normal. The assumption is reasonable because the sample size is expected to be large.

the nonresponse bias is serious for those RSA 911 variables, then we will use a more complex and time consuming procedure for nonresponse weight adjustment. The best way to utilize the rich frame data for this purpose the response propensity score methodology.

The propensity score methodology was proposed by Rosenbaum and Rubin (1983 and 1984) to draw causal inferences from observational studies (see also a review article by D'Agostino, 1998). The method has been applied to handle nonresponse in sample surveys (e.g., Little, 1986; Smith et al., 2000; Vartivarian and Little, 2003). It is assumed that given a set of covariates (denoted as A) the response indicator variable R is independent of the survey variable Y , that is, the nonresponse mechanism is missing at random (MAR). Define the conditional response propensity as

$$p(A) = \text{Prob}(R = 1 | A).$$

Then, R and Y are independent conditionally on $p(A)$. Therefore, adjusting the base sample weights by multiplying the inverse of $p(A)$ to them produce unbiased weights for Y . Since $p(A)$ is unknown, it is estimated by using logistic regression of A on response status of the full sample and the estimate (predicted value $\hat{p}(A)$ of the logistic regression) is used to adjust the sample weight. This strategy is the most effective under this approach in removing the nonresponse bias. However, it could introduce too much volatility in the adjusted weights, which would cause a large increase in variance. To avoid this, we will group the sample units into an appropriate number of nonresponse adjustment cells with units having similar propensity scores. The adjustment cells are often created based on the quintiles of the distribution of $\hat{p}(A)$. This procedure will dampen the increase in the variance but cause to leave some nonresponse bias. Nevertheless, such a compromise would be desirable from the mean squared error perspective.

Candidate variables for A include subgroup indicator that defines the design strata, disability category, significant disability indicator, demographic variables (age, gender, and race/ethnicity), geography, previous closure indicator, service duration, levels of education at application and at closure, individualized education program (IEP) indicator, employment status and earnings variables at application and closure, and public support status and amount variables at application and at closure, medical insurance at application and at closure, and type of closure at exit given in the RSA 911 file. These are considered more important for nonresponse adjustment than many other variables in the RSA 911 file not listed here. (Appendix G has the full list of variables.) Initially, we will try to use as many variables as possible in the calculation

of the propensity score with appropriate categorization and redefining new variables based on original variables. Some interaction terms may have to be included in the final model.

Once the procedure produces a new set of nonresponse adjusted weights, we will once again examine the nonresponse bias using RSA 911 variables as described earlier with CHAID-based nonresponse adjusted weights. This process may indicate that we should create more adjustment cells than quintiles. We will repeat this process until we get an acceptable set of nonresponse adjusted weights.

At around the time when this process is finished, we anticipate the SSA administrative data files linked to the RSA 911 data file will be available for further investigating the nonresponse bias issue for the earning data. These data are expected to be available after the baseline data collection, and possibly after each follow up collection. The respondent sample distributions based on the nonresponse adjusted weights would be checked against the population distributions using these administrative data sources.

Other Sources of Error

Besides the usual sampling error, there are other sources of error, of which the most important are response or measurement error resulting from the respondent's memory lapse, omission or addition, reluctance to divulge sensitive information, cognitive disability, or an insincere attitude. RSA has planned for a validation study to be conducted on the baseline reporting of earnings and Social Security disability benefits using available administrative data as the standard for comparison.

The validation study will examine the extent to which there is measurement error in interview-reported annual earnings for 2006 compared to employer-reported earnings from UI wage records for the same time period, as well as a comparison with IRS annual earnings data from the SSA, and SSA monthly reported earnings. This validation plan assumes that UI employer-reported earnings will be available from at least two states, that tests will be based on the aggregation of records across states, and that SSA will assist in making the comparison of IRS earnings (because of restrictions on access to IRS earnings data).

The tests can be run for all consumers for which there are UI wage record data. The plan is to submit SSNs for all the baseline survey respondents from a given state to the UI agency for matching against wage record data for the state. The information received from the UI

agency will be used to determine if the consumer was covered by UI and the amount of earnings for the quarters pertaining to 2006. Because only a few states are expected to provide UI wage record data for the study, the results of the validation study will not generalize to the full sample of PVRES participants. In addition, the number of cases to be studied per state could be small, depending on which states provide UI wage records and the number of sampled consumers that are covered by the UI system. In the largest states, we could expect at most about 300-400 consumers' records to be examined to verify employment and earnings.

The process for comparing 2006 annual earnings to IRS annual earnings will involve working directly with SSA to provide information from its analysis of the IRS data file. The study team will provide SSA with the SSNs for the sampled former consumers who responded to the survey and ask SSA to provide the annual reported earnings for submitted cases within specified dollar ranges. The use of ranges is required because of restrictions on the use of the IRS earnings data (not to provide exact dollar amounts). Another limitation is that not all consumers are expected to have filed a tax return for 2006.

Another validation of earnings to be conducted is to examine differences in reported earnings for the respondent's current job, or main job (job with the most hours in the past 12 months if not currently working), with earnings data from UI wage records and from SSA's TRF link file. The TRF link file is composed of nine SSA administrative files linked to the RSA 911.

Social Security benefits are transfers that reflect relatively consistent recipient status and income levels once eligibility has been established. We expect that most recipients will report receipt and dollar amount accurately. The plan is to submit all SSNs for the baseline survey respondents to SSA for matching against its data file to determine receipt and amount of benefits, particularly SSI and SSDI. The purpose is to validate the survey responses on receipt and dollar amount.

Beside the response error analysis, we will also analyze the frequency of reporting errors concerning receipt of benefits (and the amount of benefits received):

- What percentage of survey cases disagree with the record data (for example, the percentage of survey cases that report a benefit amount less than the amount in the record data)?

- How can disagreeing cases of receipt be characterized in terms of explanatory variables (use of logistic regression to examine the probability of survey response of “no,” but record data indicate “yes”)?

For this investigation, we will use a simple but useful response error model given as follows:

$$Y_i = \mu_i + E_i,$$

where, for the i -th sample respondent, Y_i is response value, μ_i is the true value, and E_i is response error, which is considered random. This means that if the respondent is asked the same question repeatedly, a different answer will be given each time. If there is no response error, the error term is zero. However, in reality this is hardly true, and we are more concerned about potentially large error for the earnings and benefits data. Assuming that UI and SSA data provide true values (μ_i), we can investigate the magnitude and variability of the response error term in the above model for the sample units for which the UI and SSA data available. Let such sample be a simple random sample with size n .

For all $i = 1, 2, \dots, n$, the mean and variance of Y_i over the repeated interviewing are given by,

$$E(Y_i) = E(\mu_i) + E(E_i) = \mu_i + \beta, \text{ where } E(E_i) = \beta.$$

$$V(Y_i) = V(E_i) = V(\varepsilon_i) = \sigma^2, \text{ where } \varepsilon_i = E_i - \beta.$$

If β is zero, there is no bias in the mean of Y_i . There are three important aspects in the error model above to be examined:

Whether the bias term β is negligible;

Whether μ_i and ε_i are correlated; and

Whether the magnitude of σ^2 is large.

The bias term can easily be estimated by $\hat{\beta} = \frac{1}{n} \sum_{i=1}^n (Y_i - \mu_i)$. We will perform a statistical test to examine whether $\beta = 0$. If it is significantly different from zero, we will calculate the relative size of the bias in terms of the mean of μ_i , namely, $\hat{\beta} / \bar{\mu}$ where $\bar{\mu} = \sum_{i=1}^n \mu_i / n$. If this ratio is large, analysis results that involves Y_i should be interpreted with caution. We will also examine the second aspect. It is important because non-zero correlation could endanger regression analysis, which involves Y_i . The third aspect is also important

because it will add extra variability to the variance of an estimate, where Y_i is involved. If the bias term β is non-negligible, then the mean squared error of an estimate, where Y_i is involved, will be underestimated because the bias term cancels out in the variance formula. This may result in misleading inferences.

Reported values collected by the survey are not compatible with administrative data, and compatible values should be constructed using related variables.

Facing a large bias and/or large error variance, we will try to find out the cause and modify the instrument to mitigate the effect of the problem as much as possible. Otherwise, RSA will consider removing these items from later waves of data collection. RSA will also consider the potential for using available administrative data in place of reported earnings and/or benefits. However, there are challenges to the use of administrative data in place of survey data:

- SSA-TRF file data is not expected to be available for years after 2007 (the survey baseline).
- There are time lags in obtaining administrative data (2009 annual earnings would not be available until December 2010, after PVRES is completed).
- UI wage record data is not expected to be available for all states covered in the survey.

B4. Tests of Procedures or Methods

A paper version of the telephone interview was pretested with former VR consumers in five states. In one state, the entire instrument was administered by telephone to four former consumers selected to provide representation of persons with mental retardation and physical impairments. In the other four states, only two or three selected sections of the instrument were administered by telephone to no more than four former consumers. This matrix design insured that no question was asked of more than nine individuals. Four categories of former consumers were sought for interviews in these four states: consumers with mental illness, consumers with mental retardation, the hard of hearing, and transition-age consumers. After respondents completed the partial telephone interview, they participated in a cognitive laboratory to discuss the survey questions. The cognitive lab facilitators prepared summary reports and telephone interviewers prepared notes of their observations. All were debriefed at the end of data collection. The interview protocol was modified on the basis of pretest findings. Changes were

made to the sequencing of questions, question wording, and the way in which question clarifications are provided for interviewers.

In addition, the contractor will provide an early assessment of the sample acquisition process in a report to be delivered six weeks after telephone contact commences. As mentioned previously, interviewers will begin calling sample participants 10 days after the advance letter is mailed. We believe that anything less than four weeks of attempting to reach participants by telephone will not allow sufficient time to provide useful information. Therefore, we plan to produce frequencies describing the status of recruiting five weeks after the advance letters are mailed and four weeks after calling begins. These will be evaluated, and we will submit a report with recommendations to RSA.

B5. Consultations on Statistical Aspects of the Design

Name	Affiliation	Telephone Number
Hyunshik Lee	Westat	301-610-5112
Frank Bennici	Westat	301-738-3608
Susan Stoddard	InfoUse	510-549-6520
Linda LeBlanc	Westat	301-251-4285

References

- D'Agostino, R.B., Jr. (1998). Propensity Score Method for Bias Reduction for the Comparison of a Treatment to a Non-randomized Control Group. *Statistics in Medicine*, 17, 2265-2281.
- Little, R.J. (1986). Survey Nonresponse Adjustments. *International Statistical Review*, 54, 139-157.
- Mitchell, S., Ciemnecki, A., and Markesich, J. (2006). Removing Barriers to Survey Participation for Persons with Disabilities. Washington, DC: Mathematica Policy Research, Inc.
- Rosenbaum, P.R., and Rubin, D.B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70, 41-55.
- Rosenbaum, P.R., and Rubin, D.B. (1984). Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. *Journal of American Statistical Association*, 79, 516-524.
- Smith, P.J., Rao, J.N.K., Battaglia, M.P., Daniels, D., and Ezzati-Rice, T. (2000). Compensating for Nonresponse Bias for in the National Immunization Survey Using Response Propensities. *Proceedings of American Statistical Association, Section on Survey Research Methods*, 641-646.
- Thornton, C., Fraker, T., Livermore, G., Stapleton, D., O'Day, B., Silva, T., Martin, E.S., Kregel, J., and Wright, D. (2006). Evaluation of the Ticket to Work Program: Implementation Experience during the Second Two Years of Operations (2003-2004). Washington, DC: Mathematica Policy Research, Inc.
- Vartivarian, S., and Little, R. (2003). On the Formation of Weighting Adjustment Cells for Unit Nonresponse. University of Michigan Department of Biostatistics Working Paper Series.