Evaluating the Within Household Selection Procedures for In-Person U.S. Adult Literacy Surveys

Tom Krenzke[1] and Lin Li[1]
[1]Westat, 1650 Research Boulevard, Rockville, MD 20850

Abstract
The 2003 National Assessment of Adult Literacy (NAAL) and the international Adult Literacy and Lifeskills (ALL) surveys each involved stratified multi-stage area sample designs. During the last stage, a household roster was constructed, the eligibility status of each individual was determined, and the selection procedure was invoked to randomly select one or two eligible persons within the household. The objective of this paper is to evaluate the within-household selection procedure used and update the procedure for future literacy surveys. The analysis is based on current household composition data and intracluster correlation coefficients using the adult literacy data. In our evaluation, several feasible household selection rules are studied, considering effects from clustering, differential sampling rates, cost per interview, and household burden. In doing so, an evaluation of within-HH sampling under a two-stage design is extended to a four-stage design and some generalizations are made to multi-stage samples with different cost ratios.

Key Words: Intracluster correlation, design effects, multi-stage sampling

1. Introduction

The 2003 National Assessment of Adult Literacy (NAAL), conducted for the National Center for Education Statistics, involved a stratified four-stage cluster design that resulted in 18,500 completed assessments administered to adults age 16 and older. In the NAAL, counties were grouped to form Primary Sampling Units (PSUs), which were stratified and selected in the first stage. In the second stage, Secondary Sampling Units (SSUs) were formed and selected within the sampled PSUs. Subsequently, households (HHs) were selected within SSUs, and one sample person (1-SP) was randomly selected for HH sizes up to 3 ($B \leq 3$), and two persons (2-SPs) were selected for HH sizes greater than 3 ($B > 3$), where B denotes the number of eligible persons per HH. This rule followed the within HH sampling approach used in the first cycle of NAAL, conducted in 1992[1]. An evaluation of the selection rule was conducted using the current US HH size distribution and intraclass correlation coefficients computed from the 2003 survey. In doing so, an evaluation of within-HH sampling under a two-stage design (Clark and Steel, 2007) is extended to a four-stage design, as used in the NAAL survey, and some generalizations are made to multi-stage samples with different cost ratios.

The data used for the evaluation includes literacy measures from three scales derived from three types of literacy -- prose, document, and quantitative[2]. Two types of estimates are used; averages (e.g., average prose literacy score) and percentage of adults at some level[3] of literacy (e.g., percentage *Below Basic* prose literacy). In addition to the NAAL data, the evaluation also uses data from the international Adult Literacy and Lifeskills (ALL) -- a multi-stage clustered sample survey that was conducted in 2003 and measured similar types of literacy[4] in the US. Table 1 provides a summary of each survey's design and structure.

---

[1] For more information on the first cycle of NAAL, refer to http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2001457 (current as of October 2008).

[2] For more information about the NAAL types of literacy, refer to http://nces.ed.gov/NAAL/fr_tasks.asp (current as of October 2008).

[3] For a discussion of the literacy levels used in NAAL, see http://nces.ed.gov/NAAL/perf_levels.asp (current as of October 2008).

[4] For more information on ALL literacy types and levels, refer to http://www.statcan.ca/english/freepub/89-603-XIE/2005001/pdf/89-603-XWE part1.pdf (current as of October 2008).

Table 1: Features of the NAAL and ALL surveys

Survey	*Area sample*	*Completes*	*Data collection*	*Assessments*	*Ages*	*Within-HH sampling rule*
NAAL	*PSUs, SSUs*	18,500	Screener	Prose	16+	$B \leq 3$, $b = 1$
	HHs, Persons		Interview	Document		$B > 3$, $b = 2$
			Assessment	Quantitative		
ALL	*PSUs, SSUs,*	3,400	Screener	Prose	16-65	$B \leq 3$, $b = 1$
	HHs, Persons		Interview	Document		$B > 3$, $b = 2$
			Assessment	Numeracy		

Note: PSU= Primary Sampling Unit, SSU = Secondary Sampling Unit, b= sample size, B = HH size

A discussion of the design considerations that helped form the evaluation of the within-HH sampling rules is provided in Section 2. Section 3 discusses the computation of intra-HH correlations under multi-stage sample designs, and focuses on incorporating the clustering impact from the initial stages of sample selection when deciding on a within-HH selection rule. An evaluation of selection rules was conducted using data from the in-person adult literacy surveys and the results are provided in Section 4.

2. Design Considerations

There are a number of factors that need to be considered when evaluating the within-HH selection rules for surveys such as NAAL and ALL. The remainder of this section will discuss the impact of the following factors on within-HH sampling: HH burden, clustering persons within HHs, differential sampling rates, multi-stage sampling, cost considerations, computerized systems, domains of interest, and HH composition.

HH burden. For the adult literacy surveys, the interview and the assessment take about an hour and a half to administer in total. Therefore, one concern about selecting more than one person per HH is the increase of burden to the HH and the impact on response rates. However, Table 2 shows there is no significant difference (0.05 significance level) in the refusal rates between 1- and 2-SP HHs in ALL and NAAL.

Table 2: Refusal rates by 1- and 2-SP HHs for the adult literacy surveys

Survey	*Subgroup*	*Refusal rate %*
NAAL	1-SP HHs	16.3
	2-SP HHs	15.7
ALL	1-SP HHs	17.6
	2-SP HHs	16.2

Clustering persons within HHs. Many surveys limit the selection to one SP per HH because of concerns over the increased clustering effect (i.e., increasing effect on variance estimates) associated with multiple SPs per household. The *DEFF* due to clustering can be expressed as: $DEFF_{clu} = 1 + (b - 1)\, Rho$, where $b = \Sigma\, (M_B / M)\, b_B$, M_B = number of HHs of size *B*, *M* = number of HHs, and b_B = sample size of persons within HHs of size *B* (Kish, 1965). This design effect (*DEFF*) component increases when the sample size within a HH increases or when the value of the intracluster correlation (*Rho*) increases. As given in Cochran (1977), *Rho* can be approximated as: $Rho = 1 - \frac{\sigma_w^2}{\sigma^2}$,

where $\sigma_w^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{i.})^2 / (n-a)$, and $\sigma^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{..})^2 / (n-1)$.

Differential sampling rates. A clustering effect is not the only factor that increases the variance. Increases in variance are also due to differential sampling rates (or, differential weights). Under a 1-SP per HH strategy, the increase is directly related to the variation in HH size – since the sampling rate could vary from 1 out of 1, to 1 out of 7 or more. The *DEFF* due to differential sampling rates is expressed as: $DEFF_{wgt} = \Sigma\, (p_B / k_B)\, \Sigma\, (p_B\, k_B)$, where, $p_B = N_B / N$,

N_B = number of eligible persons in the population in HHs of size B, N = number of eligible persons in the population, and k_B = sampling rate within HHs of size B (Kish, 1965). Under certain conditions, the overall *DEFF* can be expressed as the product of the clustering and differential sampling rate components: $DEFF = DEFF_{clu} \times DEFF_{wgt}$. Kalton, Brick, and Lê (2005) suggest this product is applicable when the weights are random or approximately random.

To arrive at a self-weighting sample, persons within HHs would need to be selected at a constant rate. However, a rate-based approach is not preferred in most surveys since it would result in walking away from a portion of single-person HHs and, thus, would increase the cost of the survey. We limit the alternative rules under consideration to those with a minimum of 1-SP per HH. Out of concern for burdening HHs, the maximum sample size was set to two. The sampling rules under consideration are:

1. Take1: 1-SP no matter the HH size.
2. Rule2: 1-SP for HH sizes up to 2; otherwise 2-SPs are selected.
3. NAAL3: 1-SP for HH sizes up to 3; otherwise 2-SPs are selected.
4. Rule4: 1-SP for HH sizes up to 4; otherwise 2-SPs are selected.
5. Frac5: take at least 1-SP, but no more than 2-SPs and the sample size is a fraction.

While the Take1 approach does not attempt to reduce the *DEFF* due to differential sampling rates, it is not subject to a clustering impact. However, the other four approaches listed above provide a reduction in the differential sampling rate component (with the Frac5 approach resulting in the most reduction) while introducing a clustering effect. Figure 1 illustrates the best options under a two-stage HH design with fixed effective sample size of persons, without any cost considerations.The US national HH size distribution from the 2007 Current Population Survey was used for this illustration. As shown in Figure 1, the fractional approach is the best rule for a wide range of values of *Rho*. The fractional approach can be programmed into a computerized system when enumerating and selecting HH members (more discussion on computerized systems follow). If computerized systems are not available for screening, then the best approach for low values of *Rho* is the more clustered approach, Rule2; and the NAAL3 rule is best for *Rho* values greater than about 0.34.

<u>Multi-stage sampling</u>. For multi-stage area designs, the clustering impact of sampling within HHs is dampened by the clustering due to PSUs and SSUs. That is, more persons within a HH can be selected for surveys with a large amount of clustering due to the first two-stages of sampling. More discussion of this distinction is provided in Section 3.

<u>Cost considerations</u>. The cost of screening a HH in a 1-SP per HH design versus the cost of interviewing/assessing a second person in a HH is investigated in an extensive analysis presented later in this paper.

<u>Computerized systems</u>. Computerized systems, such as Computer-Assisted Personal Interview (CAPI), have the capability of handling fractional sample sizes, that is, if the sample size for a HH with two eligible persons is 1.6, then two persons are selected 60 percent of the time at random, and one person is selected 40 percent of the time. Computerized systems also have the capability of sorting the list of eligible persons and selecting 2-SPs with a systematic random sample. Another benefit is that the selection program can be tested and validated prior to data collection.

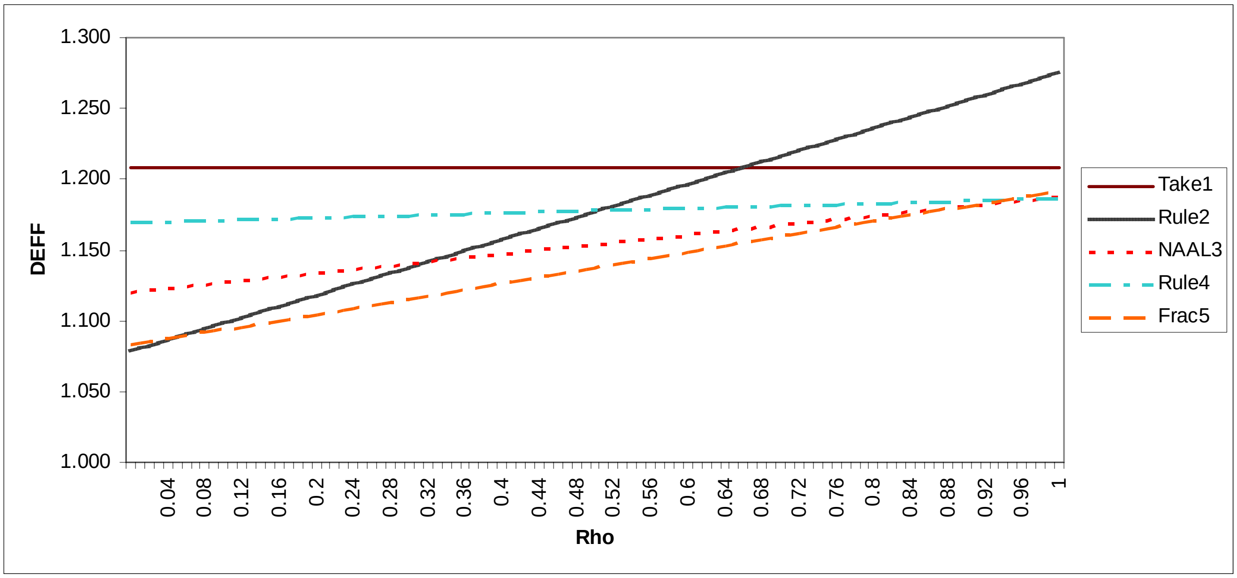

Figure 1: Initial analysis of within-HH selection rules

Domains of interest. As mentioned earlier, optimal within-HH sampling depends on the magnitude of the clustering effect associated with the variable of interest. The clustering effect may be much smaller when the variable is associated with a subgroup of the population, rather than the entire population. For example, when a key reporting domain is gender in a survey of the adult population, it is likely that the reporting category of males will have an average of 1 SP per HH, and less likely to have 2 male SPs, which would introduce a clustering effect. Therefore, when there are multiple domains of interest in a typical HH, it is often beneficial to select more than 1-SP within a HH. Refer to Mohadjer and Curtin (2008) for an example of design considerations for a survey with focus on multiple subgroups of the population.

HH composition. Lastly, one may want to consider the HH composition and relationships of persons within a HH when devising the selection rule. Table 3 displays values of *Rho* for various relationships between HH members, for HHs with 2-SPs in the NAAL survey. *Rho* varies greatly by HH member relationships. The relationships were derived from gender and age.

Table 3: *Rho* for NAAL assessment scores by HH member relationships

Estimate	*Siblings*	*Child-guardian*	*Married*	*Others*
Number of HHs with 2-SPs	111	205	180	434
Average prose score	0.42	0.35	0.70	0.59
Average document score	0.40	0.27	0.72	0.54
Average quantitative score	0.46	0.36	0.63	0.56
Percentage Below Basic prose	0.52	0.41	0.79	0.67
Percentage Below Basic document	0.54	0.40	0.78	0.60
Percentage Below Basic quantitative	0.51	0.41	0.77	0.65

3. Estimation of *Rho* under Multi-Stage Sampling

The discussion about *Rho* thus far has been related to a two-stage design, but both NAAL and ALL have four stages of sampling. The total variance can be decomposed into four between-variance terms attributable to PSUs, SSUs, HHs, and Persons, as follows: $\sigma_T^2 = \sigma_{PSU}^2 + \sigma_{SSU(PSU)}^2 + \sigma_{HH(SSU)}^2 + \sigma_{PERS(HH)}^2$.

As shown below, when applying a two-stage approach to estimate *Rho* for a four-stage sample design, the numerator not only contains the between HH component, but also contains contributions from the between PSU and between SSU components inflating the values of *Rho* for our purpose.

$$Rho = 1 - \frac{\sigma^2_{PERS(HH)}}{\sigma^2_T} = \frac{\sigma^2_{PSU} + \sigma^2_{SSU(PSU)} + \sigma^2_{HH(SSU)}}{\sigma^2_T}$$

Therefore, when evaluating rules for within-HH sampling under a multi-stage design, we assume the PSU and SSU design will be the same in the future. This can be accomplished by limiting our focus to within SSU sampling. Therefore, the computation of *Rho* is contained within SSUs, that is, it is done in a compact manner without effect from the PSU and SSU components, where the denominator is limited to the last two stages of sampling, as follows: $\sigma^2_{PERS(SSU)} = \sigma^2_{HH(SSU)} + \sigma^2_{PERS(HH)}$. We refer to this as the compact *Rho*, denoted by Rho^*, expressed as:

$$Rho^* = 1 - \frac{\sigma^2_{PERS(HH)}}{\sigma^2_{PERS(SSU)}} = \frac{\sigma^2_{HH(SSU)}}{\sigma^2_{PERS(SSU)}}, \text{ where, } \sigma^2_{PERS(HH)} = \sum_{h=1}^{SSU} \frac{M_h}{M} \sigma^2_{PERS(HH),h}, \text{ and } \sigma^2_{PERS(SSU)} = \sum_{h=1}^{SSU} \frac{M_h}{M} \sigma^2_{PERS(SSU),h}.$$

Table 4 shows the Rho^* values for average NAAL and ALL literacy assessment scores as well as values of *Rho* computed under a two-stage design assumption. When including the clustering impact from the first two stages of the four-stage design, the values of the compact Rho^* are much smaller than *Rho*. For example, the two-stage *Rho* for the NAAL average prose score is 0.57 and the compact Rho^* is equal to 0.33. The table also shows that values of the compact Rho^* for average scores are at about the same level for NAAL (range from 0.32 to 0.33) and ALL (range from 0.29 to 0.39). There is some variation by the type of estimate as well; values of Rho^* for ALL are 0.1 to 0.2 lower for the percentage in *Level 1* or *2* than for the average scores. Values of Rho^* can also vary by HH size as shown in Figure A-1 in Appendix A.

Table 4: Compact Rho^* for literacy assessment scores

	*Compact Rho**		*Two-stage Rho*	
Estimate	*NAAL*	*ALL*	*NAAL*	*ALL*
Number of HHs with 2-SPs	930	162	930	162
Average prose score	0.33	0.38	0.57	0.60
Average document score	0.33	0.29	0.53	0.50
Average quantitative/numeracy score	0.32	0.39	0.54	0.58
Percentage Below Basic(NAAL)/Level 1or 2 (ALL) prose	0.42	0.28	0.65	0.44
Percentage Below Basic (NAAL)/Level 1 or 2 (ALL) document	0.39	0.28	0.61	0.37
Percentage Below Basic quantitative (NAAL)/Level 1 or 2 (ALL) numeracy	0.40	0.17	0.62	0.36

4. Evaluation and Results

We compared the current sampling rules with optimal sampling rules by minimizing a variance-cost (*VC*) function, which is the product of the *DEFFs* due to clustering and weighting, and a cost function:

$$VC = DEFF^*_{clu} \times DEFF_{wgt} \times n(c_p + \frac{c_{HH}}{b}), \quad \text{where,} \quad DEFF^*_{clu} = \frac{k + \sum_B \frac{M_B}{M}(1 + (b_B - 1)Rho^*_B)}{k+1}, \quad k = \frac{\sigma^2_{PSU} + \sigma^2_{SSU(PSU)}}{\sigma^2_{HH(SSU)} + \sigma^2_{PERS(HH)}},$$

c_p = cost per added person, c_{HH} = cost per added HH, and Rho^*_B is computed as described in Section A1 in Appendix A.

Note that the *VC* function represents the additional cost of increasing the overall sample size to offset the increase in variance due to the *DEFF* components. For the *DEFF* due to clustering, we include a term (the *k*-value) in the numerator and the denominator in order to account for the reduction of the within-HH clustering effect because of the contributions to total variance from the PSU and SSU stages in multi-stage sampling.

As shown in Table 5, the variance ratio *k*, which is the variance from the first two stages divided by the variance from the last two stages, ranges from 0.68 to 1.61 across types of assessments and estimates for the ALL survey.

Table 5: Values of *k* for the ALL sample

ALL estimate	*k*
Average prose score	0.95
Average document score	1.56
Average quantitative/numeracy score	1.13
Percentage in Level 1 or 2 prose	0.68
Percentage in Level 1 or 2 document	1.61
Percentage in Level 1 or 2 numeracy	1.10

Table 6 provides the results for optimal integer solutions as computed by a computational algorithm developed, as described in Section A2 of Appendix A. The table shows that as the cost ratio increases from 0.5 to 1 for $k = 1$, we would want to take more persons per HH, that is, 2 out of 2 instead of 1 out of 2. As the variance ratio goes from 1 to 3, the only change is for HH size of 2 and cost ratio of 0.5. That is, when variance ratio is equal to 3, it is beneficial to take 2 out of 2, instead of 1 out of 2.

Table 6: Optimal solutions

		Integers				*Fractional*				*Walk-away*			
k	C_{HH} / C_p	$B = 1$	$B = 2$	$B = 3$	$B = 4$	$B = 1$	$B = 2$	$B = 3$	$B = 4$	$B = 1$	$B = 2$	$B = 3$	$B = 4$
1	0.5	1	1	2	2	1	1.4	2	2	0.6	1.3	2	2
1	1	1	2	2	2	1	1.6	2	2	0.9	1.6	2	2
1	2	1	2	2	2	1	1.9	2	2	1	1.9	2	2
3	0.5	1	2	2	2	1	1.6	2	2	0.8	1.5	2	2
3	1	1	2	2	2	1	1.8	2	2	1	1.8	2	2
3	2	1	2	2	2	1	2	2	2	1	2	2	2

Table 6 also gives the results when fractional sample sizes are allowed. The variance and cost ratios for NAAL and ALL tend to be about 1, where it appears that selecting 1 out of 1, 1.6 out of 2, and 2 otherwise is the best rule. The effects of cost and variance ratios are clearer under the fractional sample sizes when compared to the integer solutions.

If the cost of conducting a screener is small in relation to the cost of interviewing, then variances can be reduced using the fractional walk-away approach. Table 6 shows optimal walk-away sample sizes. Under this approach, for example, a sample size of 0.9 indicates that we walk away from 10 percent of the HHs where $B = 1$. If the cost of screening is a very small portion of the cost of interviewing, then the optimal design may involve walking away from many more HHs.

Under the likely NAAL/ALL parameters for cost ratios ($C_{HH} / C_p = 1$) and variance ratios ($k = 1$), when compared to the Take1 approach, the *VC* function can be reduced by about 9 percent by using the NAAL/ALL sampling rule, 19 percent by using the optimal integer solution, 20.4 percent using the optimal fractional solution, and 20.6 using the optimal walk-away approach. In general, the gains from deviating from the Take1 approach grow as the cost per additional HH (i.e., screening) increases. The average cluster sizes for each approach are given in Table 7. For the NAAL and optimal integer rule, the average cluster size indicates the percentage of HHs with 2 SPs. For example about 6 percent of the HHs would have 2-SPs under the NAAL3 strategy.

Table 7: Percent reduction of NAAL3 and optimal solutions from Take1 strategy and average cluster sizes

		Percentage reduction from Take1 strategy				*Average cluster sizes*			
k	C_{HH} / C_p	*NAAL3*	*Integer*	*Fractional*	*Walk-away*	*NAAL3*	*Integer*	*Fractional*	*Walk-away*
1	0.5	8.2	13.0	15.8	18.0	1.06	1.18	1.38	1.21
1	1	9.1	19.2	20.4	20.6	1.06	1.68	1.48	1.45
1	2	9.9	26.1	26.1	26.1	1.06	1.68	1.63	1.63
3	0.5	8.6	17.3	18.7	19.0	1.06	1.68	1.48	1.37
3	1	9.5	23.7	23.9	23.9	1.06	1.68	1.58	1.58
3	2	10.4	30.2	30.2	30.2	1.06	1.68	1.68	1.68

Lastly, a sensitivity analysis was conducted by varying the values of Rho^*. A regression model was fit on the percentage reduction from the Take1 strategy of the *VC* function, with the independent variables being the approach (NAAL3, integer, fractional, walk-away), cost ratio (0.1, 0.5, 1, 2, 10), variance ratio (1, 3, 5) and Rho^* (+/- 0.1). For the range of data, it was somewhat surprising to see that Rho^* did not have a significant impact (p-value = 0.106) on the percentage reduction of the *VC* function, while the other factors were significant.

Summary

Several design considerations were taken into account when evaluating the within-HH selection rule for the NAAL and ALL surveys, including taking into account clustering effects from initial stages of sampling. To facilitate the evaluation, we formulate a way to incorporate PSU and SSU variance contributions into the computation of the *DEFF* due to clustering and the intra-HH correlation when deciding how many persons and how many HHs to select in a multi-stage sample design. In addition, a computational algorithm was developed to compute optimal sample size solutions, incorporating the *DEFFs* due to clustering, differential sampling rates, and costs.

In general, the main factors on the percentage reduction of the *VC* function from the Take1 approach are the level of dominance from the PSU and SSU variance components in multi-stage sampling, the cost ratio, and the rule used. For the range of data evaluated, Rho^* was not a significant factor on the reduction in *VC* from the Take1 approach. In general, NAAL rule improves on the widely-used Take1 approach. The optimal integer rule improves on NAAL rule. However, the optimal fractional rule has limited gains over optimal integer rule. The optimal walk-away rule has gains over other rules for lower cost ratios. Lastly, when the first two variance components dominate and cost ratio is high, then the integer, fractional and walk-away rules are essentially the same.

Acknowledgements

The authors acknowledge valuable contributions from Leyla Mohadjer, Bob Fay and Keith Rust.

References

Clark, R., and Steel, D. (2007), Sampling within Households in Household Surveys, *Journal of the Royal Statistical Society*, Series A, *170*, 63-82.
Cochran (1977). *Sampling Techniques*. 3rd ed. New York: Wiley.
Kalton, Brick, and Lê (2005). Estimating Components of Design Effects for Use in Sample Design, *Household Sample Surveys in Developing and Transition Countries*, Chapter VI, United Nations, New York.
Kish (1965). *Survey Sampling*. New York: Wiley.
Mohadjer, L. and Curtin, L. (2008). Balancing sample design goals for the National Health and Nutrition Examination Survey. *Survey Methodology*, *34*(1)

Appendix A

A-1 Estimates of *Rho** by HH size

Survey estimates are not attainable for *Rho** by HH size since only 1 SP was selected for HH size of 3 or less, and since the sample size was too small to create estimates for each HH size of 4 or more. Therefore, estimates of *Rho** by HH size are modeled using Census data. Figure A-1 shows *Rho** on the y-axis and HH size on the x-axis. The upper line is from the Census PUMS education attainment for ages 25+. The upper line shows that education attainment is more similar among HHs with two adults, perhaps more likely to be married couples. It shows a drop off when going from two to three adults. We captured the variation in HH size by computing the ratio of *Rho** for the NAAL prose literacy scores to the *Rho* for the Census PUMS education attainment among HHs with $B > 3$, and applying the ratio to the PUMS *Rho* across all HH sizes. The resulting values are the estimates of compact Rho_B^*, for $B = 1, 2, \ldots 11$.

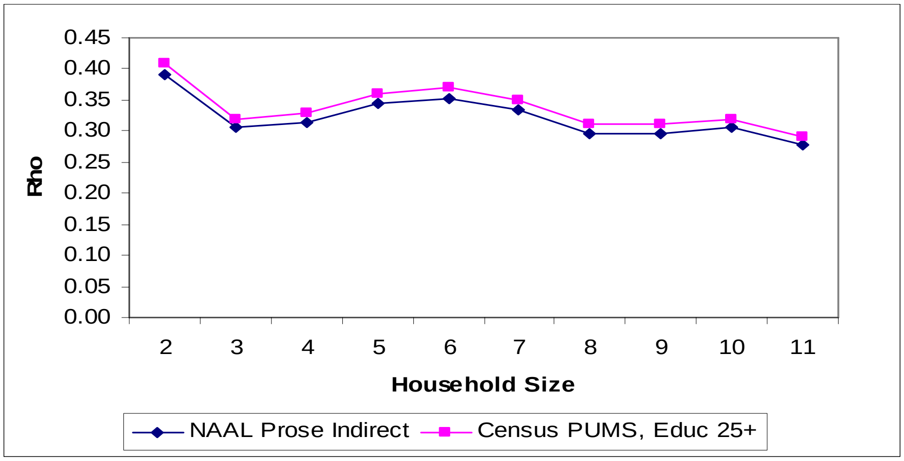

Figure A-1: Estimates of *Rho** for NAAL by HH size

A-2. Computational Algorithm

A computational algorithm was developed to arrive at optimal within-HH sample sizes for each HH size *B*. The algorithm was constructed to generate optimal integer or fractional solutions that capture the effects of clustering, differential sampling rates and cost, under the constraints of at least one selected per HH and no more than 2. Here are the steps of the algorithm (all processing runs converged within four iterations):

- Initialize by setting $b = 1$ for all values of B (Take1).
- Compute $DEFF^*_{clu}$, $DEFF_{wgt}$, c_p, c_{HH}, and $VC(0)$.
- Do $I = 1$ to 5.
    - Do $B = 1$ to 11.
        - Compute $DEFF^*_{clu}$, $DEFF_{wgt}$, c_p, c_{HH}, and VC for all $1 \le b_B \le 2$, given the set of b_B, for all $B' \neq B$.
        - Identify the b_B with the smallest value of VC.
    - End.
    - if $VC(I) = VC(I\text{-}1)$ then stop.
- End.