



International
Labour
Organization

ILO-IPEC Interactive Sampling Tools No. 5

Listing the sample Primary Sampling Units (PSUs)

Version 1

December 2014

**International
Programme on
the Elimination
of Child Labour
(IPEC)**

**Fundamental Principles and Rights at Work (FPRW) Branch
Governance and Tripartism Department**

Copyright © International Labour Organization 2014
First published 2014

Publications of the International Labour Office enjoy copyright under Protocol 2 of the Universal Copyright Convention. Nevertheless, short excerpts from them may be reproduced without authorization, on condition that the source is indicated. For rights of reproduction or translation, application should be made to ILO Publications (Rights and Permissions), International Labour Office, CH-1211 Geneva 22, Switzerland, or by email: pubdroit@ilo.org. The International Labour Office welcomes such applications.

Libraries, institutions and other users registered with reproduction rights organizations may make copies in accordance with the licences issued to them for this purpose. Visit www.ifrro.org to find the reproduction rights organization in your country.

ILO-IPEC

ILO-IPEC Interactive Sampling Tools No. 5 - Listing the sample Primary Sampling Units (PSUs) / International Labour Office, International Programme on the Elimination of Child Labour (IPEC) - Geneva: ILO, 2014

ACKNOWLEDGEMENTS

This publication was elaborated by Mr. Farhad Mehran, consultant, for ILO-IPEC and coordinated by Mr. Federico Blanco Allais from IPEC Geneva Office.

Funding for this ILO publication was provided by the United States Department of Labor (Projects GLO/13/21/USA & GLO/10/55/USA).

This publication does not necessarily reflect the views or policies of the United States Department of Labor, nor does mention of trade names, commercial products, or organizations imply endorsement by the United States Government.

The designations employed in ILO publications, which are in conformity with United Nations practice, and the presentation of material therein do not imply the expression of any opinion whatsoever on the part of the International Labour Office concerning the legal status of any country, area or territory or of its authorities, or concerning the delimitation of its frontiers.

The responsibility for opinions expressed in signed articles, studies and other contributions rests solely with their authors, and publication does not constitute an endorsement by the International Labour Office of the opinions expressed in them.

Reference to names of firms and commercial products and processes does not imply their endorsement by the International Labour Office, and any failure to mention a particular firm, commercial product or process is not a sign of disapproval.

ILO publications and electronic products can be obtained through major booksellers or ILO local offices in many countries, or direct from ILO Publications, International Labour Office, CH-1211 Geneva 22, Switzerland. Catalogues or lists of new publications are available free of charge from the above address, or by email: pubvente@ilo.org or visit our website: www.ilo.org/publns.

Visit our website: www.ilo.org/ipec

Available in electronic PDF format only.

Photocomposed by ILO-IPEC Geneva.

1. Introduction

This document describes the use of the template “PSU listing” of the SIMPOC Interactive Sampling Tools. After selecting the sample PSUs, the next step in conventional two-stage sample designs is to freshly list the households in the sample PSUs before drawing the final sample of households for interviewing. Listing means identifying households currently living in an area and recording their basic characteristics.

The main purpose of listing is to up-date the sampling frame and in particular the secondary sampling units or households within the primary sampling units, taking into account population movements and new household formations that have occurred since the last preparation of the sampling frame. It aims at covering the newly constructed buildings with living quarters and taking into account demolished or vacant buildings, or transformed dwellings no longer used as living quarters, such as dwelling addresses turned to stores or workshops, or living quarters used as secondary housing units or for summer holidays. The purpose of listing is thus to ensure that every household currently residing in the sample PSUs will have a non-zero probability of selection. Indirectly, listing also helps to improve the population projection of the country and therefore the accuracy of the sampling weights of the survey.

Listing may also be used to redirect the sample design for targeting the sample units of interest. For example, in a child labour survey, the listing operations may be used to identify households with children and limit the sample selection to these households. In some cases, the listing operations may further be used to obtain an initial broad estimate of the phenomenon of interest. For example, during the listing operations the number of children not attending school may be identified in each listed household, and the results used to estimate the number of children not attending school as a broad proxy of working children.

The listing template of the SIMPOC interactive sampling tools provide for recording the households listed in a child labour survey and select samples of households by systematic sampling with equal probabilities under different stratification and allocation schemes depending on the extent of information recorded in the listing operations. Section 2 describes the input part of the listing template with numerical illustrations. Section 3 describes the output part and Section 4 the part on intermediate calculations. Finally, Section 5 presents results on the efficiency of alternative designs described in the output template.

2. Input values

Input values are the information to be entered in the template by users. There are two types of input values: a few input parameters specifying the design of the second stage sampling of households and a set of input data feeding the template with the data obtained from the listing operations. The following table gives a numerical illustration of the input template.

The input parameters are shown in the top panel. There are three input parameters: *b* is the sample-size specifying the number of sample households to be drawn in each PSU (in this example *b* is equal to 15); *alpha C* is the parameter to be used for allocating the listed households into sub-strata within the PSU (here *alpha C* is equal to 1 corresponding to proportional allocation); finally the parameter *random* in the random number between 0 and 1 specifying the random start for selecting the sample households within PSUs).

INPUT VALUES									
Input parameters									
b = 15			alpha C = 1			random = 0.318383			
Input data									
Stratum or domain	Sample PSU id no.	Buiding serial no.	Dwelling serial no.	Household serial no.	Name	Address	A	B	C
				208			847	149	17
Stratum 1	PSU 1	1	1	1	AA BB	CC DDDD	2	1	1
Stratum 1	PSU 1	2	2	2	AA BB	CC DDDD	4	1	1
Stratum 1	PSU 1	3	3	3	AA BB	CC DDDD	7	1	1
Stratum 1	PSU 1	3	3	4	AA BB	CC DDDD	2	1	1
Stratum 1	PSU 1	4	4	5	AA BB	CC DDDD	5	1	1
Stratum 1	PSU 1	5	5	6	AA BB	CC DDDD	6	1	1
Stratum 1	PSU 1	5	6	7	AA BB	CC DDDD	5	1	0
Stratum 1	PSU 1	6	7	8	AA BB	CC DDDD	3	1	0
Stratum 1	PSU 1	7	8	9	AA BB	CC DDDD	1	1	0
Stratum 1	PSU 1	7	8	10	AA BB	CC DDDD	7	1	0
Stratum 1	PSU 1	8	9	11	AA BB	CC DDDD	7	1	0
Stratum 1	PSU 1	9	10	12	AA BB	CC DDDD	1	1	0
Stratum 1	PSU 1	10	11	13	AA BB	CC DDDD	4	1	0
Stratum 1	PSU 1	10	12	14	AA BB	CC DDDD	6	1	0
Stratum 1	PSU 1	10	13	15	AA BB	CC DDDD	2	1	0
Stratum 1	PSU 1	10	14	16	AA BB	CC DDDD	1	1	0
Stratum 1	PSU 1	10	15	17	AA BB	CC DDDD	6	1	0

The input data are given in the lower panel of the template. The input data are in rectangular form with 10 columns and as many rows as there are households listed in the listing operations. The input values are entered by stratum and by PSU and within PSUs sorted by the C variable described later.

In this example, there are 208 listed households of which the first 17 are shown in the table. The first column specifies the stratum. It can be in the form of a code or a text such as the name of the region or the administrative district of the country. The second column specifies the PSU generally in the form of multi-digit identification code. The input data in this example concern one stratum (Stratum 1) and three PSUs (PSU 1, PSU 2, and PSU 3). Here only part of the data for PSU 1 is seen on the table.

The next five columns specify the household, starting with the building serial number, then the dwelling serial number, next the household serial number and finally the name and address of the head or the reference person of the household. In the numerical illustration seen on the table, the listing has covered 10 buildings, 15 dwellings and 17 households. Households 3 and 4 live in the same dwelling (dwelling number 3, building number 3) where there is a mother with a child is renting a room in a house where a seven-member household lives. Similarly, households number 9 and 10 live in the same dwelling (dwelling number 8, building number 7) where a young working student is renting a room in a house also occupied by a seven-member household.

In addition to buildings with serial numbers 3 and 7 occupied by multi-households in the same dwelling, there is building number 5 (a villa) and building number 10 (an apartment house) occupied respectively by two and five households in separate dwellings. The remaining buildings are occupied each by a single household.

The serial numbers of buildings, dwellings, and households and the information on the name and address of the household are actually not used in the calculations of the outputs. They are here for the identification of households and reflect the contents of typical listing forms. For the calculations, the only requirement is that each row represents one distinct household sorted by stratum and within stratum by sample PSU.

The last three columns of input data are for recording the data on household characteristics collected during the listing operations. There is provision for up to three data items. They are labeled in generic form as A, B and C because the data items collected may differ depending on the contents of the listing form. In the present numerical illustration, the three data items refer to:

- A. Number of household members
- B. Presence of any child 5 to 17 years old in the household (B=1 if there is at least one child; B=0 otherwise)
- C. Presence of any working child in the household (C=1 if there is at least one working child; C=0 otherwise)

In the numerical illustration, household number 1 in PSU 1 has 2 household members. At least one is a child 5 to 17 years old and working. Household number 2 has 4 members. There is at least one child 5 to 17 years old in the household. There is also at least one child 5 to 17 years old in the household who is working. As

indicated in the top row with bold lines, there are altogether 847 household members in the 208 households covered by the listing operations, 149 households have at least one child 5 to 17 years old and 17 households have at least one working child in the household.

Other examples of specific data items that could be collected in a child labour survey would be:

- A. Number of children (5 to 17 years old) in the household
- B. Number of children (5 to 17 years old) not attending school in the household
- C. Number of working children (5 to 17 years old) in the household

In the context of a general household survey or a labour force survey, the data items may be simpler such as

- A. Number of household members
- B. Number of male members
- C. Number of female members

In practice, many other possibilities may be considered for the data items to be collected during listing.

3. Output values

The output values of the templates are in the form of three sets of column data giving the sample selection of households with their probabilities of selection based on three alternative sample designs:

Design 1. Sampling all listed households

Design 2. Sampling only households with children

Design 3. Stratified sampling of households with children

The output values are calculated automatically based on the input values. A numerical illustration corresponding to the input values presented earlier is shown in the following table, followed by a description under each design.

OUTPUT VALUES								
Design 1			Design 2			Design 3		
Sampling all listed households			Sampling only households with children (5-17 years old)			Stratified sampling of households with children and with working children		
Strata A	Probability HSLD	Sample A	Strata B	Probability HSLD	Sample B	Strata C	Probability HSLD	Sample C
29	3	42	37	4	37	37	17	37
1	20.5%		1	29.4%		2	33.3%	1
1	20.5%	1	1	29.4%	1	2	33.3%	
1	20.5%		1	29.4%		2	33.3%	
1	20.5%		1	29.4%		2	33.3%	1
1	20.5%		1	29.4%	1	2	33.3%	
1	20.5%		1	29.4%		2	33.3%	
1	20.5%	1	1	29.4%		1	28.9%	
1	20.5%		1	29.4%	1	1	28.9%	1
1	20.5%		1	29.4%		1	28.9%	
1	20.5%		1	29.4%		1	28.9%	
1	20.5%		1	29.4%		1	28.9%	1
1	20.5%	1	1	29.4%	1	1	28.9%	
1	20.5%		1	29.4%		1	28.9%	
1	20.5%		1	29.4%		1	28.9%	
1	20.5%		1	29.4%	1	1	28.9%	1
1	20.5%		1	29.4%		1	28.9%	
1	20.5%	1	1	29.4%		1	28.9%	

Design 1 - Sampling all listed households

The first three columns of the output template give the results of sample selection under design 1. Design 1 refers to the conventional sampling scheme under which all listed households, whether they have children or not, are sampled, using systematic sampling with equal probabilities within each PSU. The first column gives the stratum number. Under design 1 there is only one stratum. Therefore the value in this column is always 1. The second column gives the probability of selection of each household in the sample and the third column indicates whether the household has been selected in the sample (value 1) or not (value 0).

For example, in row 1 corresponding to household 1 in PSU 1, the probability of selection is

$$Probability_{HSLD_1} = \frac{15}{73} = 0.205$$

because the total number of listed households in PSU 1 is 73 among which b=15 are selected under the design. The third column of row 1 is blank or 0 indicating that the household has not been selected in the sample.

The distribution of sample households by PSU and type of household under Design 1 is given below.

**Design 1. Distribution of sample households by PSU and type of household
Numerical illustration based on input and output values of template**

Sample PSU	Total	Number of sample households			
		Without children	With children	With no working children	With working children
PSU 1	15	4	11	10	1
PSU 2	12	5	7	7	0
PSU 3	15	4	11	9	2
Total	42	13	29	26	3

Of the 15 sample households selected in PSU 1, 4 are without children and 11 with children. Among those with children, 10 households have no working children and one has working children. PSU 2 has altogether 12 households listed, less than the sample take of 15. Therefore, all households in the PSU are selected in the sample, which means the sample in PSU 2 is in fact a census and each household has selection probability equal to 1. Five of the households have no children and 7 have children, among which none has working children. In PSU 3 with a total of 123 households listed, 15 sample households could be selected, 4 were without children and 11 with children, and among them 9 with no working children and 2 with working children.

Under design 1, the total sample size in 42 households composed of 13 households without children and 29 with children, of which 3 households with working children. The total number of sample households with children, the total number of households with working children and the total sample size are respectively given in the columns of the top row of the output template.

Design 2 - Sampling only households with children

Under design 2, sample selection is also systematic sampling with equal probability but restricted only to households with children (5 to 17 years old) identified during the listing operations. Under design 2 there are therefore in effect two strata specified in the first column: Stratum 1, household with children and Stratum 0, households with no children. The probabilities of selection are given in the next column. For example, for household 1 in PSU 1 identified in the first row, the probability of selection is calculated as

$$Probability_{HSLD_1} = \frac{15}{51} = 0.294$$

where the numerator is the sample-size $b=15$ and the denominator the number of households with children listed in PSU 1. The third column of row 1 is blank indicating that the household has also not been selected in the sample under design 2.

The distribution of sample households by PSU and type of household under design 2 is given below. In this case, the total sample size is 37 households. In PSU 1 the number of sample households is 15, all with children because of the design. Two of these households have working children. In PSU 2 the number of listed households with children is 7 and all are selected in the sample because the number is less than the sample-size ($b=15$). None of these households have working children. In PSU 3 the number of sample households is 15, all with children because of the design. Two of these households have working children.

**Design 2. Distribution of sample households by PSU and type of household
Numerical illustration based on input and output values of template**

Sample PSU	Total	Number of sample households			
		Without children	With children	With no working children	With working children
PSU 1	15	0	15	13	2
PSU 2	7	0	7	7	0
PSU 3	15	0	15	13	2
Total	37	0	37	33	4

The total sample size under design 2 is 37 households. All 37 are households with children (8 more than under design 1). Four of the sample households under design 2 have working children (one more than under design 1).

Design 3 - Stratified sampling of households with children

Under design 3, the households with children are first stratified within each sample PSU according to the information obtained during listing on whether they have working children or not (item C of the listing form). The sample-size of $b=15$ households is then allocated among two strata according to the general power allocation, that is in proportion to

$$n^\alpha$$

where α is the allocation parameter with $\alpha=1$ corresponding to proportional allocation, $\alpha=0$ to equal allocation, $\alpha=1/2$ to square-root allocation, etc. Negative power allocation may also be considered, for example, $\alpha=-1$ corresponding to the harmonic or reverse allocation.

In the numerical illustration presented earlier, the allocation parameter specified in the input template is $\alpha=1$, i.e., proportional allocation. Accordingly, in the output template, the sample-take in each PSU is allocated in proportion to the number of listed households with working children (Stratum 2) and number of listed households with children but not working (Stratum 1). Then, the sample households are selected by systematic sampling with equal probability within each stratum. The results are shown in the last three columns of the output template.

For example, household 1 in PSU 1 shown in the first row is in stratum 2 under design 3 because it has been identified as a household with working children during listing. Its probability of selection is calculated as

$$Probability_{HSLD_1} = \frac{2}{6} = 0.333$$

where the numerator is the number of sample households allocated to stratum 2 under the proportional allocation specified by the allocation parameter $\alpha=1$ and the denominator is the number of households with working children in the stratum identified during the listing operations. The resulting probability of selection is recorded in the second column of the design 3 portion of the output template. The third and last column of the template is a dichotomous variable indicating whether the household has been selected in the sample or not. In the case of household 1 in PSU 1, the household has been selected in the sample and the indicator value is recorded as 1.

The overall distribution of sample households by PSU and type of household under design 3 is given in the top panel of the table below. The results happen to be identical to those under design 2. The total sample size is 37 households. All 37 are households with children. Four of the sample households under design 2 have working children. If the households were allocated by inverse proportional allocation ($\alpha=-1$), all 17 households with working children identified during listing would be selected in the sample as shown in the bottom panel of the table.

**Design 3. Distribution of sample households by PSU and type of household
Numerical illustration based on input and output values of template**

Proportional allocation of households among strata ($\alpha=1$)					
Sample PSU	Total	Number of sample households			
		Without children	With children	With no working children	With working children
PSU 1	15	0	15	13	2
PSU 2	7	0	7	7	0
PSU 3	15	0	15	13	2
Total	37	0	37	33	4
Inverse proportional allocation of households among strata ($\alpha=-1$)					
PSU 1	15	0	15	9	6
PSU 2	7	0	7	7	0
PSU 3	15	0	15	4	11
Total	37	0	37	20	17

4. Intermediary calculations

The derivation of the probabilities of selection and the sample selection itself are carried out as intermediary calculations in a separate part of the template. They are described below for each of the three designs. Under design 1, in the first column the number of households listed in the PSU is calculated for each listed household (N). This becomes the denominator of the probability of selection of the household calculated in the second column. The numerator is the sample-take (b=15) specified in the input template or more precisely min(b, N) to take account of very small PSUs like PSU 2 not shown here where the number of listed households is less than the sample-take. The next two columns are for sample selection using the algorithm described in the template on selection of PSUs with unequal probabilities except that here the selection is for households and with equal probabilities.

INTERMEDIARY CALCULATIONS														
Design 1				Design 2					Design 3					
Number of A-hslds in PSU	Selection probability of A-HSLD	random	sample A	Stratum B	Number of B-hslds in PSU	Selection probability of B-HSLD	random	sample B	Stratum C	Number of hslds in PSU in C-strata	Allocation of sample-take b	Selection probability of C-HSLD	random	sample C
	42	-0.318383	42			37	-0.318383	37					-0.318383	37
73	20.5%	-0.112904	0	1	51	29.4%	-0.024265	0	2	6	6	100.0%	0.681617	1
73	20.5%	0.092576	1	1	51	29.4%	0.269852	1	2	6	6	100.0%	1.681617	1
73	20.5%	0.298055	0	1	51	29.4%	0.563970	0	2	6	6	100.0%	2.681617	1
73	20.5%	0.503535	0	1	51	29.4%	0.858088	0	2	6	6	100.0%	3.681617	1
73	20.5%	0.709014	0	1	51	29.4%	1.152205	1	2	6	6	100.0%	4.681617	1
73	20.5%	0.914494	0	1	51	29.4%	1.446323	0	2	6	6	100.0%	5.681617	1
73	20.5%	1.119973	1	1	51	29.4%	1.740441	0	1	45	9	20.0%	5.881617	0
73	20.5%	1.325453	0	1	51	29.4%	2.034558	1	1	45	9	20.0%	6.081617	1
73	20.5%	1.530932	0	1	51	29.4%	2.328676	0	1	45	9	20.0%	6.281617	0
73	20.5%	1.736412	0	1	51	29.4%	2.622793	0	1	45	9	20.0%	6.481617	0
73	20.5%	1.941891	0	1	51	29.4%	2.916911	0	1	45	9	20.0%	6.681617	0
73	20.5%	2.147370	1	1	51	29.4%	3.211029	1	1	45	9	20.0%	6.881617	0
73	20.5%	2.352850	0	1	51	29.4%	3.505146	0	1	45	9	20.0%	7.081617	1
73	20.5%	2.558329	0	1	51	29.4%	3.799264	0	1	45	9	20.0%	7.281617	0
73	20.5%	2.763809	0	1	51	29.4%	4.093382	1	1	45	9	20.0%	7.481617	0
73	20.5%	2.969288	0	1	51	29.4%	4.387499	0	1	45	9	20.0%	7.681617	0
73	20.5%	3.174768	1	1	51	29.4%	4.681617	0	1	45	9	20.0%	7.881617	0

Under design 2, the same procedure is carried out except that in first column the stratum of the household is determined i.e., whether the household has children (stratum 1) or is without children (stratum 0). For households without children, sample selection is not carried out as under design 2, sampling is only carried out among households.

Under design 3, like in design 2, the stratum of the household is first determined based on the listing information of the input template. Households without children are in stratum 0, households with children but not working are in stratum 1 and households with working children are in stratum 2. In the second column the sample-take of the PSU ($b=15$) is allocated to the stratum of the household, and in the third column the total number of households in the stratum for that PSU is calculated. The ratio of the second column to the third calculated in the next column is the probability of selection of the household. Finally, in the last two columns of the template, the sample selection of the household is made using the sample algorithm as under design 1 or 2.

5. Efficiency of alternative designs

The efficiency of a design may be measured in terms of its design effect, defined as the ratio of the variance of the estimator of the main variable calculated under the design to the variance of the estimate that would have been obtained under simple random sampling with the same size.

$$deff = \frac{Var_d(Y)}{Var_o(Y)}$$

where $Var_d(Y)$ is the variance of the estimator under the proposed design and $Var_o(Y)$ is the variance of the estimator that would have been obtained under a reference design. The design effect expresses how well the one design compares to the reference design. The lower the design effect is the higher the efficiency of the proposed design.

In the present context, $Var_d(Y)$ is the variance of the estimator of the total number of children in child labour under design 2 (or design 3) and $Var_o(Y)$ is the corresponding variance under design 1. Under all designs (1, 2 and 3), households are selected according to systematic sampling with equal probability. So the variances are essentially the same as those of simple random sampling. The difference is the population over which the sample is drawn. Under design 1, the population is all households listed within a sample PSU. Thus,

$$Var_o(Y) = N^2(1-f) \frac{\sigma_y^2}{n}$$

where N is the total number of households listed in the sample PSU, n is the sample size, $f = n/N$ the sampling rate and σ_y^2 the variance of the number of children in child labour in a household.

Under design 2, households are selected among households with children. Therefore, the variance is expressed as

$$Var_d(Y) = N_d^2(1 - f_d) \frac{\sigma_{y_d}^2}{n}$$

where N_d is the total number of households with children identified during listing in the sample PSU, n is the sample size (the same as under design 1), $f_d = n/N_d$ is the sampling rate and $\sigma_{y_d}^2$ the variance of the number of children in child labour in a household with children.

To evaluate these variances, it is necessary to calculate σ_y^2 and $\sigma_{y_d}^2$. In general, let y be the number of children in a sample household,

$$y = \sum_{j=0}^{j=m} X_j$$

where X_j is a dichotomous variable indicating whether child j is in child labour ($X_j=1$) or not ($X_j=0$), and m is the random variable indicating the number of children in the household. The number of children may be zero ($m=0$), in which case $y=0$. In the case of design 2 (and design 3), sampling is over households with children. Therefore,

$$y_d = \sum_{j=1}^{j=m} X_j$$

Expressions of y and y_d are sums of random numbers of random variables. Accordingly, assuming m has a Poisson distribution with parameter λ (the average number of children per household) and X_j a Bernoulli distribution with parameter p (the child labour rate), it can be shown that y has both a Poisson distribution with parameter λp .¹ Thus, the mean and variance of y are equal,

$$\sigma_y^2 = \lambda p$$

In the case of y_d the expression of the variance is more complicated as the random sum starts with $j=1$. It can however be derived that

$$\sigma_{y_d}^2 = \frac{(1 - e^{-\lambda p})(\lambda p + \lambda^2 p^2) - \lambda^2 p^2}{(1 - e^{-\lambda p})^2}$$

Replacing these expressions of σ_y^2 and $\sigma_{y_d}^2$ in the expressions of $Var_o(Y)$ and $Var_d(Y)$ presented earlier, the efficiency of design 2 in relation to design 1 may be calculated in terms of the parameters N , N_d , λ and p . (N = number of households in a sample PSU, N_d = number of households with children in the PSU, λ = average number of children per household in the PSU, and p = predicted overall child labour rate).

¹ Sum of random number of random variables,

http://www-ma4.upc.edu/~jfabrega/pipe/sum_random-e-handout-4pp.pdf

These results have been obtained under the assumption that there is no correlation among the random variables X_j 's within households, i.e., the work statuses of children in a given household are independent of each other. To cover the situation when the work statuses of children in households may be correlated, similar results have been obtained, not presented here, using a constant correlation coefficient within households (ρ). A value $\rho=0$ means no correlation and a value $\rho=1$ means full correlation, i.e., if a child of a household works, all children of the household also work, and vice versa if a child does not work, all don't.

The following table presents the results for design 2 (relative to design 1) of the numerical illustration described earlier. The results show that design 2 (sampling only households with children identified during listing) is more efficient than the conventional design 1 (sampling all listed households irrespective of presence or absence of children) in all circumstances. The gain in standard error ranges from about 30% to more than 50%. In general, the lower is the child labour rate, the higher the gain in standard error of the estimate. The gain in standard error is also higher, the lower the average number of children per household, and the lower the correlation of work status among children in the same household.

Efficiency of design 2 (sampling only households with children identified during listing)

Average gain in standard error of estimate of percentage of children in child labour

Numerical illustration: Three PSUs with 208 households, 149 households with children

% children in child labour (p)	No correlation within household ($\rho=0$)					Full correlation within household ($\rho=1$)				
	Average number of children per hsltd (λ)					Average number of children per hsltd (λ)				
	0.1	0.5	1	1.5	2	0.1	0.5	1	1.5	2
1%	50.2%	50.2%	50.1%	50.1%	50.0%	50.2%	50.2%	50.1%	50.0%	49.9%
10%	50.1%	49.8%	49.4%	49.0%	48.6%	50.2%	50.0%	49.4%	48.4%	47.0%
20%	50.0%	49.4%	48.6%	47.8%	47.0%	50.2%	49.8%	48.6%	46.6%	44.1%
30%	50.0%	49.0%	47.8%	46.6%	45.5%	50.2%	49.6%	47.8%	45.0%	41.5%
40%	49.9%	48.6%	47.0%	45.5%	44.1%	50.2%	49.4%	47.0%	43.4%	39.2%
50%	49.8%	48.2%	46.2%	44.4%	42.7%	50.2%	49.2%	46.2%	41.9%	37.2%
60%	49.7%	47.8%	45.5%	43.4%	41.5%	50.2%	49.0%	45.5%	40.6%	35.6%
70%	49.6%	47.4%	44.8%	42.4%	40.3%	50.1%	48.8%	44.8%	39.3%	34.2%
80%	49.5%	47.0%	44.1%	41.5%	39.2%	50.1%	48.6%	44.1%	38.2%	33.2%
90%	49.5%	46.6%	43.4%	40.6%	38.2%	50.1%	48.4%	43.4%	37.1%	32.3%
100%	49.4%	46.2%	42.7%	39.7%	37.2%	50.1%	48.2%	42.7%	36.2%	31.6%

The calculations are programmed in columns AM to AV of the template and can be performed by specifying the parameters in the upper three rows of columns AN-AO. Corresponding results for evaluating the efficiency of design 3 relative to design 1 (or design 2) are more complex and have not been done.