Test LOT Quality Assurance Sampling Method in Estimating Communes with Improved Latrine in Vietnam

Nguyen, Cuong

19 July 2017

TEST LOT QUALITY ASSURANCE SAMPLING METHOD IN ESTIMATING COMMUNES WITH IMPROVED LATRINE IN VIETNAM

Nguyen Viet Cuong

Abstract

In this study, we will simulate different ways of select households according to the LC-LQAS method, and examine the accuracy of these ways to measure the percentage of communes having the sanitation coverage of 70%. We tried different sampling methods which differ in the number of clusters (village) and the number of households selected in each communes. The results suggest that the methods identify communes with the sanitation rate of at least 70% quite well. More than 90% of communes are correctly identified. The number of clusters selected in a commune plays an important role in reducing the mean squared error and increasing the correct prediction rates. Possible, two clusters within a commune with the total number of sampled households of 19 is the best choice.

Keywords: sampling method, LC-LQAS method, household survey, sanitation, Vietnam.

JEL classification: C13; C81; C83.

* Mekong Development Research Institute and National Economics University, Hanoi, Vietnam.

Email: cuongwur@gmail.com

1. INTRODUCTION

Although Vietnam has experienced poverty reduction during the decades, there are still a large proportion of population lacking access to sanitation. According the 2014 VHLSS, 20% of households did not have access to improved latrines, while this figure for ethnic minorities was much higher, at 59%. Being aware of the important role of sanitation, the Ministry of Planning and Investment (MPI) has requested the World Bank (IDA) to finance a Program-for-Results (PforR) to increase the access to improved latrines in the 21 provinces of the northern and central regions, which have the lowest sanitation coverage. This Program-for-Results, hereinafter referred to as "the Program", is expected to run from 2016 to 2020 and will cover 680 communes in 21 provinces. The provinces that will be covered under the Program are: Dak Lak, Dak Nong, Gia Lai, Kon Tum, Lam Dong, Bac Can, Bac Giang, Cao Bang, Dien Bien, Ha Giang, Hoa Binh, Lai Chau, Lang Son, Lao Cai, Phu Tho, Son La, Thai Nguyen, Tuyen Quang, Yen Bai, Binh Thuan and Ninh Thuan.

The Program will focus specifically on: (a) improving the effectiveness of the existing efforts of the government to scale up access to sanitation and improved hygiene practices; and (b) improving the sustainability of the infrastructure put into place, for both water supply and institutional sanitation facilities. For sanitation, the Program will aim to achieve Commune-Wide Sanitation (CWS) across a significant proportion of the rural communes in the target provinces. CWS requires 70 percent of households have an improved latrine.

To monitor and evaluate the program, we have to measure the main outcome of sanitation coverage at the commune level. Since the program covers a large number of communes, collection of data that are representative at the commune is very costly. Thus in

this report, we consider the Lot Quality Assurance Sampling (LQAS) method to examine whether communes have achieved the coverage of 70%. Instead of collecting representative data to estimate the coverage rate, the LQAS collects only a small number of households per commune, say less than 20 households, to predict whether the coverage rate below or larger than 70% (Lemeshow and Taber, 1991; Robertson and Valadez, 2006). In Vietnam, communes are relatively large, and collection of data on randomly selected households remains costly. Recently, Hedt et al. (2008) discussed Large Country-Lot Quality Assurance Sampling (LC-LQAS). According to this method, instead of selecting households randomly within a commune, we can select households from a certain number of clusters within the commune to reduce travelling cost.

In this study, we will simulate different ways of select households according to the LC-LQAS method, and examine the accuracy of these ways to measure the percentage of communes having the sanitation coverage of 70%.

2. Data and simulation methods

We use the data on improved latrines from the 2009 Population and Housing Census and the 2014 Intercensal Population and Housing Survey. Nguyen and Kov (2017) use small area estimation to estimate the improved data and construct household-level data that are representative at the commune level for the year 2014 (see Appendix for definition of improved latrines). In this study, we will use this data set as to estimate the coverage rate of sanitation at the commune level. These estimates are considered as the population parameter or actual rates of improved latrines in communes. Then we will draw different random samples of households from each commune, and compare the sanitation coverage

rate predicted from these samples with the population parameters. We examine the method in 21 project provinces (including 3,209 communes).

The optimal number of observations per unit is around 19. In this study, we also consider random selection of 19 households per commune or per village. Villages are considers as clusters within communes. The average number of villages per commune is around 7. There are 9 ways to select households as follows:

- Representative sample: We select a sample that is representative at the commune level. The sample size varies according to the total population of communes. On average, around 191 households are selected per commune. This method is regarded as the benchmark.
- 19 households selected within commune: 19 households are randomly selected within a commune (no cluster).
- 19 households selected within one village: one village is randomly selected from the list of villages in a commune, and then 19 households are randomly selected from the selected village (one cluster).
- 19 households selected within two villages: two villages are randomly selected within a commune, and then 9 and 10 households are selected from these two villages (two clusters).
- 19 households selected within three villages: three villages are randomly selected within a commune, and then 6, 6 and 7 households are selected from these three villages (three clusters).
- 38 households selected within one village: one village is randomly selected from the

list of villages in a commune, and then all 38 households are randomly selected from the selected village (one cluster).

- 38 households selected within two villages: two villages are randomly selected within a commune, and then 19 households are selected from each selected village (two clusters).
- 57 households selected within one village: one village is randomly selected from the list of villages in a commune, and then all 57 households are randomly selected from the selected village (one cluster).
- 57 households selected within three villages: three villages are randomly selected within a commune, and then 19 households are selected from each selected village (three clusters).

3. Simulation Results

As mentioned, there are 9 sampling methods. For each sampling method, we conduct 200 simulations. For each simulation, we compute the two following variables in each commune:

- Correct prediction: A dummy variable which is equal to 1 if the predicted rate and the actual rate of improved latrines are both $< 70\%$, or both $\geq 70\%$.
- Squared error: the square of the difference between the predicted rate and the actual rate of improved latrines.

For each sampling method, we conduct 200 simulations for 3,209 communes. It means that there are 200 simulations for each commune. We estimate the average of the four above variables across 200 simulations. Thus for each commune, we estimate the two following parameters:

- The percentage of correct prediction.
- Mean squared error.

For example, a given commune is supposed to have the rate of improved latrines of 75%. The percentage of correct prediction is estimated at 90%. It means that among 200 simulations, there are 180 simulations in which the sample estimates correctly predict the commune to have the sanitation rate equal or larger than 75%. For each simulation, we also compute the square of the difference between 75% and the estimate of the sanitation rate. The mean squared error is the average of these differences over 200 simulations.

Table 1 reports the results of the two parameters from the 9 sampling methods. It presents the average and standard deviation of the estimates across communes. This table reports the average and standard deviation for the whole 21 provinces. Tables in Appendix present the estimates for each province.

Table 1: Assessment of the sampling methods

Sampling methods	Mean squared error		The percentage of correct prediction	
	Mean	Std. Dev.	Mean	Std. Dev.
Benchmark: representative sample	3.1	3.7	99.0	5.7
19 hhs selected within a commune	53.9	50.0	96.0	10.3
19 hhs selected within one village	327.5	450.8	90.4	16.5
19 hhs selected within two villages	198.3	265.7	92.9	14.4
19 hhs selected within three villages	156.1	207.2	93.8	14.2
38 hhs selected within one village	211.9	282.2	92.3	14.1
38 hhs selected within two villages	183.1	257.9	93.1	14.6

Sampling methods	Mean squared error		The percentage of correct prediction	
	Mean	Std. Dev.	Mean	Std. Dev.
57 hhs selected within one village	162.2	209.5	93.3	13.0
57 hhs selected within three villages	126.9	184.2	94.3	14.2

Figure 1 presents the percentage of correct prediction in 21 provinces. The figure is the average of the percentage of correct prediction across 3,209 communes in the data set. The correct prediction rate is larger than 90 in all the sampling methods. The representative sample size (the benchmark) yields a very high correct prediction rate, at 99%. Random selecting 19 households in one village per commune have the lowest correction rate at 90.4%. The standard deviation does not differ largely for different sampling methods.

Figure 1: The percentage of correct prediction

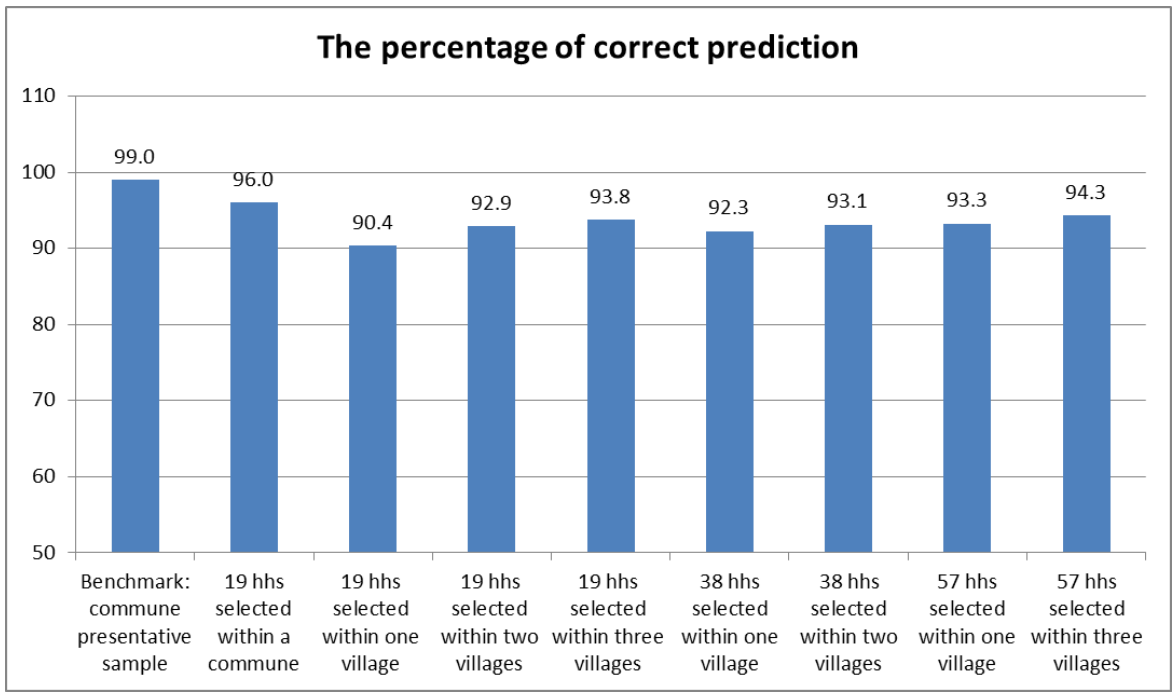
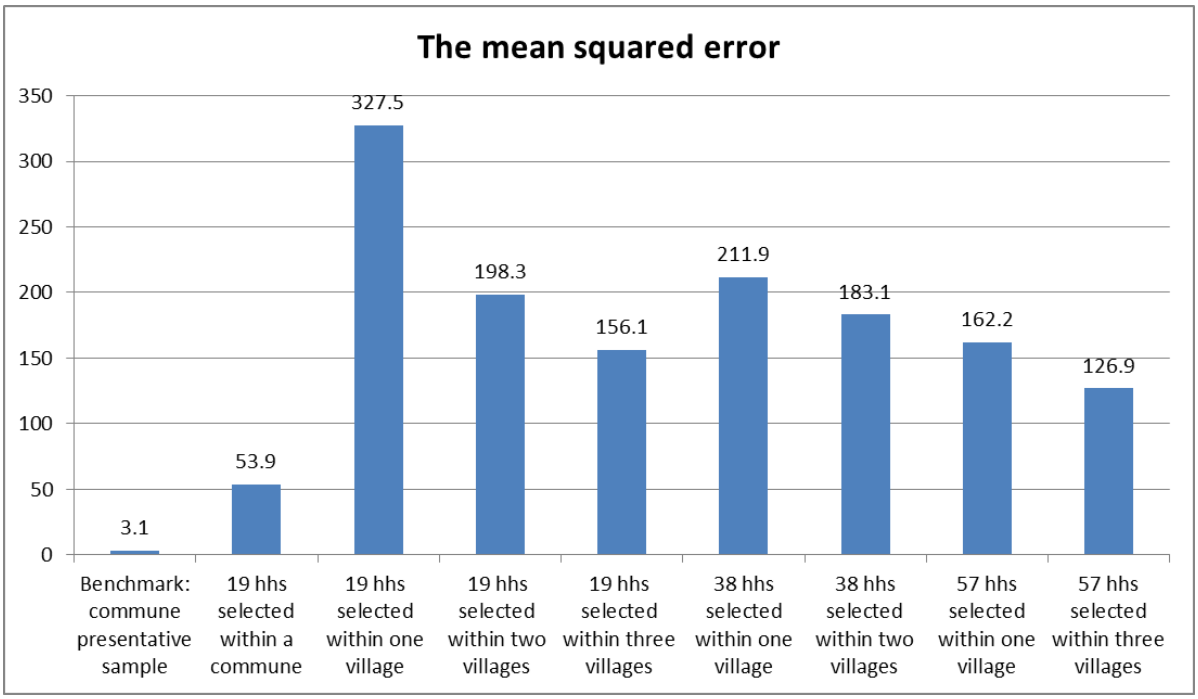


Figure 2 presents the mean squared error (MSE). High mean squared error means less accurate methods in measuring the percentage of sanitation rate. It shows clearly that

increasing the number of clusters reduce significantly the MSE. Sampling 19 households in three villages is better than sampling 38 households in two villages or 57 households in one village.

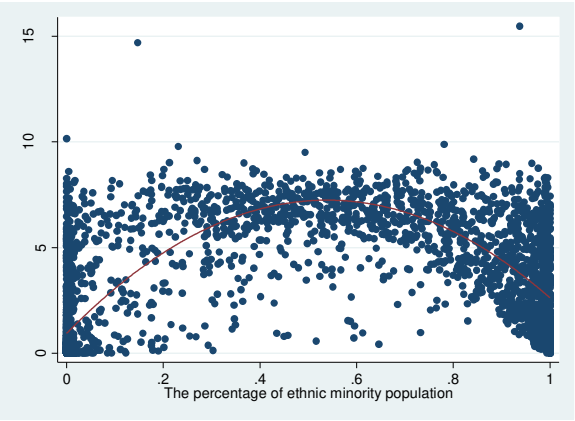
Figure 2: Mean squared error



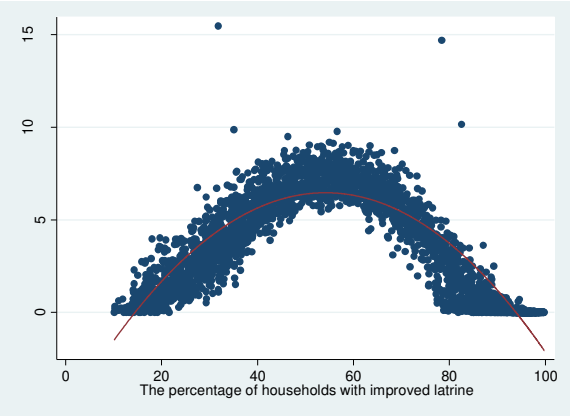
In Figure 3, we examine the relation between the mean squared error and the share of ethnic minority population and the sanitation rate in communes. It shows a clear invert U shaped relation between the mean squares error and the share of ethnic minority population as well as the rate of improved latrines. Communes with low or high share of ethnic minority population have the lowest mean squares error. Similarly, communes with low or high rates of improved latrines have the lowest mean squares error.

Figure 3: The representative sample: mean square error and ethnic minority share and sanitation rate

Mean square error and share of ethnic minority



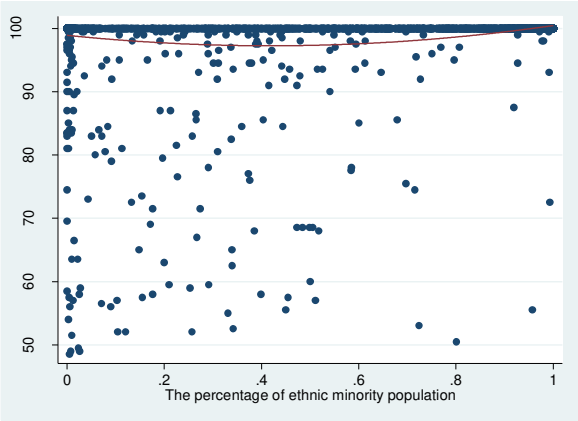
Mean square error and sanitation rate



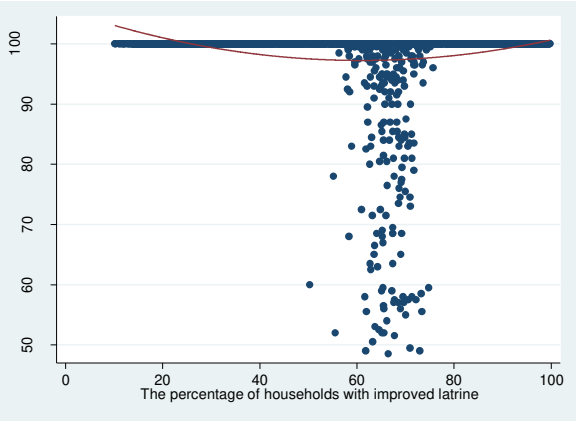
Regarding the correct prediction rate, there is a positive correlation between the correction prediction rate and the share of ethnic minority population, and a negative relation between the correction prediction rate and the sanitation rate. However, in the Figure 4 this relation is not clear.

Figure 4: The representative sample: the correct prediction rate mean square error and ethnic minority share and sanitation rate

The correct prediction rate and share of ethnic minority



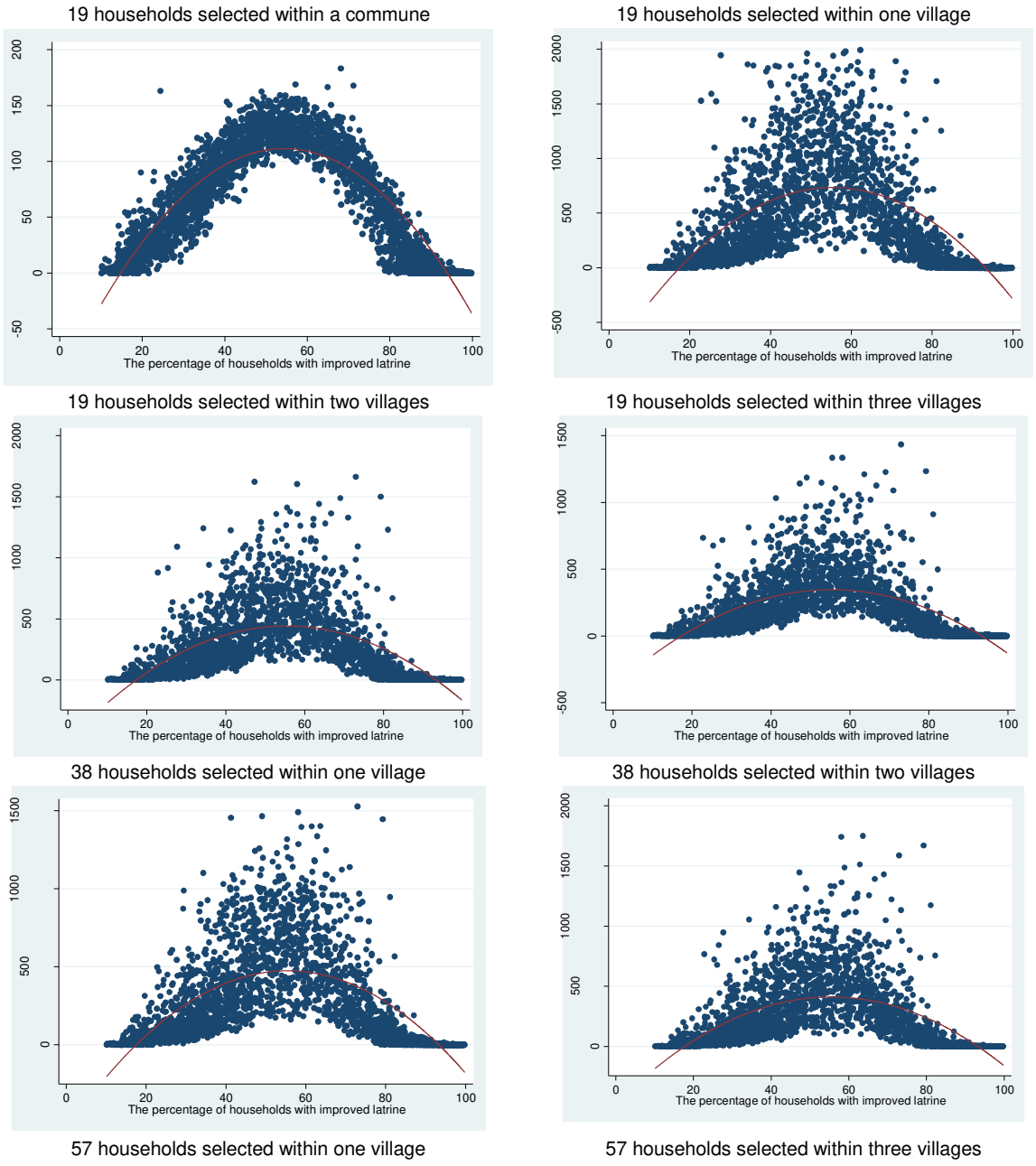
The correct prediction rate and the sanitation rate

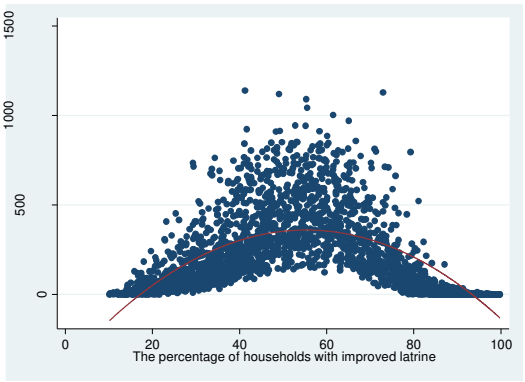


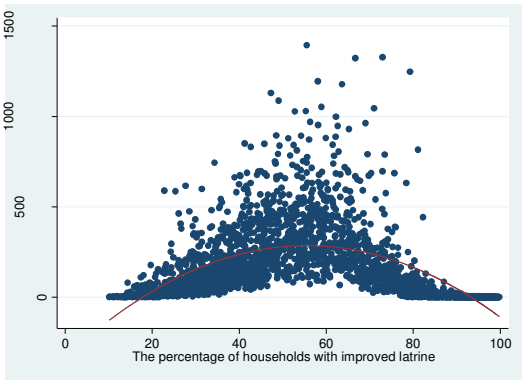
For other sampling methods, the relation between the mean squared error, the correct prediction rate, the share of ethnic minority population, and the sanitation rates are similar to the representative sampling method. In the following figures, we focus on the relation between the mean squared errors and the sanitation rate of different sampling

methods. It shows invert U shaped relation between the mean squared error and the sanitation rate of communes.

Figure 5: the mean squared error and the sanitation rate of communes.







4. Conclusions

In this study, we will simulate different ways of select households according to the LC-LQAS method, and examine the accuracy of these ways to measure the percentage of communes having the sanitation coverage of 70%. We tried different sampling methods which differ in the number of clusters (village) and the number of households selected in each communes. The results suggest that the methods identify communes with the sanitation rate of at least 70% quite well. More than 90% of communes are correctly identified. The number of clusters selected in a commune plays an important role in reducing the mean squared error and increasing the correct prediction rates. Possible, two clusters within a commune with the total number of sampled households of 19 is the best choice.

References

Elbers, C., Lanjouw, J. and Lanjouw, P., 2002, Micro-Level estimation of welfare. Policy Research Working Paper No. WPS 2911. The World Bank.

Elbers, C., Lanjouw, J. and Lanjouw, P., 2003, Micro-level estimation of poverty and inequality, *Econometrica,* 71(1), pp. 355-364.

Nguyen, Viet Cuong and Phyrum Kov (2017), Using the Small Area Estimation Method to Inform Targeting for Results-Based Scaling Up Rural Sanitation and Water Supply Program, Research report, Hanoi, Vietnam.

Hedt, B.L., Olives, C., Pagano, M. and Valadez, J.J., 2008. Large country-lot quality assurance sampling: a new method for rapid monitoring and evaluation of health, nutrition and population programs at sub-national levels.

Robertson, S.E. and Valadez, J.J., 2006. Global review of health care surveys using lot quality assurance sampling (LQAS), 1984–2004. *Social science & medicine*, *63*(6), pp.1648-1660.

Lemeshow, S. and Taber, S., 1991. Lot quality assurance sampling: single-and double-sampling plans.

Appendix

Table A.1: Definition of improved latrines

Type of latrines		Percent
Improved latrines	Septic/semi-septic tank	63.27
	Flush/soak pit	4.18
	Improved toilet with vent	0.38
Unimproved latrines	Double septic tank	11.68
	Barrel/pot	0.29
	Fishing bridge	7.10
	Others	8.73
	None/Bush/On the field	4.36
	Total	100

Table A.2. Results from representative sampling (benchmark) and sampling of 19 households

Provinces	Representative sampling (benchmark)				19 households selected within commune				19 households selected within one village			
	Mean squared error		The percentage of correct prediction		Mean squared error		The percentage of correct prediction		Mean squared error		The percentage of correct prediction	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
TP. Ha Noi	0.03	0.18	100.00	0.00	0.43	2.97	100.00	0.00	0.72	4.75	99.99	0.07
Ha Giang	3.70	9.11	100.00	0.04	46.82	41.64	99.46	2.87	231.77	334.63	96.97	8.75
Cao Bang	3.58	2.92	99.88	1.16	59.71	39.00	99.05	5.23	217.80	253.98	97.89	7.40
Bac Kan	5.46	1.53	99.99	0.11	98.79	28.66	98.01	5.69	560.87	367.03	88.06	16.10
Tuyen Quang	5.19	2.42	97.64	7.30	90.35	38.66	90.91	14.47	496.87	412.88	81.08	20.51
Lao Cai	4.55	5.09	99.60	3.63	70.89	44.48	96.38	8.89	492.95	512.03	89.27	15.79
Dien Bien	2.79	2.67	100.00	0.00	47.64	41.46	99.11	3.81	308.76	419.88	94.21	13.00
Lai Chau	3.16	8.85	99.51	4.12	37.94	35.56	99.41	4.89	191.12	334.01	97.79	7.06
Son La	3.88	2.44	99.68	2.86	69.20	44.22	98.40	6.04	462.80	504.07	91.85	14.37
Yen Bai	4.61	2.44	99.12	3.23	81.58	43.29	93.36	12.52	439.86	417.57	85.29	17.96
Hoa Binh	5.46	2.19	98.44	6.33	98.84	38.47	93.27	11.99	691.36	501.52	76.89	20.16
Thai Nguyen	4.25	2.77	98.00	8.28	75.21	48.44	93.27	12.91	411.23	414.55	84.40	17.81
Lang Son	4.34	2.06	99.92	0.73	76.20	35.49	98.96	4.18	400.63	454.89	96.03	9.39
Quang Ninh	2.04	2.74	98.41	7.57	35.27	47.03	94.28	11.75	160.09	247.07	89.94	15.80
Bac Giang	1.99	3.40	99.19	4.57	32.68	43.02	96.24	10.03	219.43	434.03	91.04	18.30
Phu Tho	3.31	2.43	99.17	3.88	59.03	43.28	94.07	11.42	266.91	335.65	88.81	16.15
Vinh Phuc	0.12	0.47	100.00	0.00	2.02	7.85	99.97	0.19	3.47	15.02	99.95	0.31
Bac Ninh	0.00	0.02	100.00	0.00	0.02	0.15	100.00	0.00	0.11	0.77	100.00	0.00
Hung Yen	0.00	0.01	100.00	0.00	0.01	0.11	100.00	0.00	0.09	1.04	100.00	0.00
Ha Nam	0.05	0.09	100.00	0.00	0.97	1.64	100.00	0.00	4.00	10.65	99.96	0.33
Thanh Hoa	2.45	2.69	99.14	5.02	43.42	47.72	96.70	9.23	264.02	399.77	90.76	17.35
Ninh Thuan	4.15	1.91	97.85	9.21	75.25	34.34	92.16	12.54	515.36	468.88	83.77	16.52
Binh Thuan	3.63	1.77	98.07	6.32	65.73	33.39	88.24	15.13	280.27	313.16	84.34	17.08
Kon Tum	3.58	2.45	99.84	1.27	63.23	43.20	98.14	6.32	438.89	506.69	91.32	14.88
Gia Lai	5.33	2.81	97.45	9.65	91.81	37.40	93.65	13.16	636.86	506.26	83.56	17.68
Dak Lak	6.07	1.72	96.42	10.96	107.64	29.79	89.15	15.61	1022.49	616.98	78.93	17.15
Dak Nong	6.25	1.61	97.71	7.50	110.34	26.81	90.35	13.86	578.78	340.03	77.16	19.39
Lam Dong	5.86	1.75	92.97	13.84	103.38	31.39	83.47	17.87	548.11	387.31	67.56	22.76

Table A.3. Results from sampling of 19 households and 38 households

Provinces	19 households selected within two villages				19 households selected within three villages				38 households selected within one village			
	Mean squared error		The percentage of correct prediction		Mean squared error		The percentage of correct prediction		Mean squared error		The percentage of correct prediction	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
TP. Ha Noi	0.54	3.56	100.00	0.00	0.51	3.28	100.00	0.00	0.57	3.75	100.00	0.03
Ha Giang	139.14	181.65	98.18	6.52	107.54	134.14	98.70	5.12	155.28	201.85	98.11	5.99
Cao Bang	139.62	148.65	98.70	6.34	115.49	118.86	98.85	6.35	148.19	161.62	98.40	6.25
Bac Kan	334.04	193.32	92.82	13.03	259.91	144.00	94.27	12.08	318.48	189.88	93.79	11.17
Tuyen Quang	297.18	219.66	84.62	18.71	230.86	156.28	86.19	18.16	325.31	260.44	84.08	18.54
Lao Cai	288.39	297.96	92.35	13.36	219.40	214.07	93.32	13.59	300.96	293.65	91.63	13.24
Dien Bien	175.27	224.80	96.90	9.54	133.42	168.11	97.54	8.56	199.27	253.55	96.13	10.21
Lai Chau	111.36	177.50	98.81	5.87	88.13	135.06	98.93	6.28	135.24	231.90	98.25	6.76
Son La	268.75	285.89	95.38	11.26	206.31	213.83	96.48	10.82	301.00	320.31	94.44	11.69
Yen Bai	263.40	229.59	88.85	16.56	202.14	171.69	90.42	16.42	279.01	248.07	88.67	15.58
Hoa Binh	414.65	291.48	83.41	17.85	327.59	230.84	84.97	18.57	417.29	280.53	83.50	15.82
Thai Nguyen	241.78	235.80	88.72	14.93	184.88	177.27	90.13	14.51	253.76	247.86	87.80	15.24
Lang Son	256.26	285.78	98.15	5.05	211.17	230.39	98.96	3.17	269.97	287.58	96.75	7.88
Quang Ninh	96.66	144.34	91.63	14.77	75.19	110.75	92.29	14.49	96.48	141.51	90.81	15.42
Bac Giang	133.10	258.69	92.39	16.24	106.21	206.39	93.07	16.31	146.29	280.11	92.73	14.48
Phu Tho	171.03	202.10	92.16	14.48	136.56	153.06	93.18	14.39	176.49	213.64	90.77	13.86
Vinh Phuc	2.67	11.09	99.97	0.18	2.33	9.69	99.99	0.05	3.27	13.85	99.93	0.50
Bac Ninh	0.12	0.79	100.00	0.00	0.10	0.64	100.00	0.00	0.06	0.40	100.00	0.00
Hung Yen	0.05	0.53	100.00	0.00	0.03	0.39	100.00	0.00	0.04	0.42	100.00	0.00
Ha Nam	3.22	6.87	99.98	0.16	3.05	6.41	100.00	0.00	2.72	5.69	99.99	0.05
Thanh Hoa	159.57	235.12	92.85	14.87	123.54	177.17	93.46	14.77	164.83	250.53	92.82	14.17
Ninh Thuan	317.05	265.14	88.05	14.32	244.52	197.44	89.95	14.23	311.91	255.27	86.57	13.65
Binh Thuan	174.86	172.08	87.03	16.26	142.94	133.13	87.92	16.64	203.25	216.50	85.08	15.76
Kon Tum	268.95	307.15	94.22	13.53	207.39	235.48	95.09	13.31	291.58	315.07	93.75	11.80
Gia Lai	381.69	292.41	89.02	15.52	297.68	221.76	90.84	14.84	426.12	334.90	86.42	16.68
Dak Lak	634.39	384.28	85.03	20.87	512.65	322.32	86.75	22.75	668.99	370.46	80.54	16.45
Dak Nong	352.46	179.60	81.49	20.79	284.30	145.48	82.18	22.43	397.45	195.17	81.20	17.24
Lam Dong	330.35	216.33	73.05	21.07	262.60	162.56	75.21	20.88	376.39	253.58	72.18	20.96

Table A.4. Results from sampling of 38 households and 57 households

Provinces	38 households selected within two villages				57 households selected within one village				57 households selected within three villages			
	Mean squared error		The percentage of correct prediction		Mean squared error		The percentage of correct prediction		Mean squared error		The percentage of correct prediction	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
TP. Ha Noi	0.38	2.62	100.00	0.00	0.52	3.60	100.00	0.03	0.29	2.07	100.00	0.00
Ha Giang	127.92	178.08	98.36	5.66	118.51	147.53	98.46	5.79	84.14	112.76	98.96	4.30
Cao Bang	123.26	145.88	98.84	6.09	120.21	119.20	98.54	5.77	85.83	95.84	99.03	6.28
Bac Kan	300.16	195.07	93.19	13.22	241.93	132.80	95.06	9.98	199.82	126.89	94.39	13.50
Tuyen Quang	262.55	210.43	85.04	18.81	260.61	204.48	85.56	17.83	175.88	143.22	87.22	18.34
Lao Cai	268.28	278.71	92.47	13.70	229.21	221.63	92.89	12.05	181.22	203.69	94.01	12.45
Dien Bien	167.60	226.28	96.71	9.88	158.02	192.97	97.12	8.47	113.14	152.10	97.52	8.37
Lai Chau	102.22	173.25	98.88	4.71	106.83	176.16	98.69	5.53	67.28	109.81	99.27	4.54
Son La	254.20	282.54	95.17	11.65	229.09	235.83	95.66	10.04	168.73	188.29	96.62	10.53
Yen Bai	242.15	235.69	89.11	17.01	218.45	186.05	90.16	14.68	166.57	163.93	90.82	17.23
Hoa Binh	376.87	277.86	84.16	18.47	311.51	206.55	86.03	15.37	266.00	207.76	86.06	19.24
Thai Nguyen	225.42	231.83	88.85	15.41	196.65	186.88	89.65	13.81	149.84	157.51	90.78	14.92
Lang Son	236.83	290.31	98.15	4.91	202.79	192.35	97.37	6.83	167.51	206.11	99.08	2.65
Quang Ninh	85.71	129.24	91.91	14.99	75.56	109.67	92.38	13.34	56.63	88.27	93.22	13.95
Bac Giang	125.02	252.17	92.80	16.06	110.29	202.15	93.61	13.21	88.02	179.66	93.63	16.26
Phu Tho	156.94	198.51	92.59	14.94	133.74	153.90	92.13	12.58	108.42	133.19	94.39	14.63
Vinh Phuc	2.21	9.30	99.99	0.05	3.01	12.36	99.97	0.24	1.57	7.01	100.00	0.00
Bac Ninh	0.06	0.39	100.00	0.00	0.05	0.35	100.00	0.00	0.06	0.46	100.00	0.00
Hung Yen	0.03	0.30	100.00	0.00	0.03	0.38	100.00	0.00	0.02	0.19	100.00	0.00
Ha Nam	2.32	5.03	100.00	0.00	2.00	4.17	99.99	0.05	1.85	4.43	99.99	0.05
Thanh Hoa	144.34	222.79	92.96	15.03	122.59	180.34	93.85	12.79	98.24	152.55	94.03	14.56
Ninh Thuan	299.27	267.35	88.49	15.53	232.28	182.73	87.78	12.96	212.82	196.84	91.85	13.46
Binh Thuan	160.18	180.28	88.39	17.13	166.62	166.28	84.99	15.70	112.28	122.53	90.68	17.10
Kon Tum	242.17	283.28	94.80	12.03	219.51	239.12	95.10	10.66	164.43	197.14	95.68	12.77
Gia Lai	346.37	278.37	89.30	15.55	331.70	256.70	88.28	15.84	240.70	201.84	91.79	14.75
Dak Lak	615.63	399.48	84.90	21.04	490.24	262.91	82.49	16.55	458.63	317.44	87.01	23.49
Dak Nong	329.43	194.53	81.30	20.56	308.60	143.61	83.26	16.70	222.95	140.95	83.87	21.78
Lam Dong	300.75	216.64	72.93	23.09	303.68	198.31	74.90	19.49	206.64	154.81	75.84	22.57