Application of Consecutive Sampling Technique in a Clinical Survey for an Ordered Population: Does it Generate Accurate Statistics?

Mohamad Adam Bujang[1]
Clinical Research Centre, Sarawak General Hospital, Ministry of Health, Malaysia

Tg Mohd Ikhwan Tg Abu Bakar Sidik
Department of Pharmacology, Faculty of Medicine, Universiti Kebangsaan Malaysia

Nadiah Sa'at
Institute for Clinical Research, Ministry of Health, Malaysia

ABSTRACT

This study aims to compare the statistical generalizations which are inferred from samples obtained by both systematic sampling and consecutive sampling and then compare both their results with the true population parameters of the target population. This study was conducted using two approaches. The first approach was a comparison between sample statistics and population parameters based on a simulation analysis to estimate the population parameters from three types of statistical distributions (i.e. Normal, Exponential, and Poisson) by using seven sub-samples and 1000 iterations. The second approach was a comparison between sample statistics and population parameters based on real-life data sets which comprise six sub-samples and four parameters. Based on results from the simulation analysis, systematic sampling offers a greater advantage by having a smaller value of mean square error (MSE) in 40 out of 70 comparisons (57.1%) while consecutive sampling has a smaller value of MSE in 29 out of 70 comparisons (41.4%). There is only one MSE comparison that was identical between systematic sampling and consecutive sampling. Based on a validation approach, systematic sampling produced more accurate statistics than consecutive sampling with six out of eight comparisons. In summary, systematic sampling offers a better advantage in terms of accuracy. However, consecutive sampling is still able to generate valid and accurate statistics despite the fact that it is a type of non-probability sampling, especially if a sufficiently large sample size has been obtained for statistical analysis. Therefore, it is recommended that in any situation when it can be difficult to apply a systematic sampling technique for a particular clinical setting, researchers may opt to apply the consecutive technique in the recruitment process as an alternative, with a limitation on making generalizations about the target population.

Keywords: *population parameters; sample statistics; systematic sampling.*

[1] **Address correspondence to Mohamad Adam Bujang:** adam@crc.gov.my

1. Introduction

In the medical field, patients are usually arranged in an ordered sequence such as a consecutive list of patients obtaining their appointments for seeking medical treatment. There is usually no sampling frame available since both the total population size and the entire population list are sometimes unknown and unspecified. Therefore, in an observational study, patient recruitment is mostly conducted by using consecutive sampling (Flamaing et al. 2015; Jensen et al. 2015; Weigner et al., 2015). It has already been known that consecutive sampling is typically far better than convenience sampling in controlling and minimizing the risk of sampling bias (Polit and Beck, 2010). Convenience sampling involves a recruitment procedure in which all respondents will be selected for inclusion in the sample merely by virtue of being conveniently available to the researcher. On the other hand, consecutive sampling shall necessitate a researcher to recruit all respondents on a first-come and first-served basis so long as the subjects fulfill all the eligibility requirements stipulated by both the inclusion and exclusion criteria (Bowers et al., 2011; Bujang, 2017).

Consecutive sampling is a type of non-probability sampling method which is commonly adopted as the sampling technique for an ordered population, particularly in a clinical survey. Consecutive sampling is defined as a sampling method that obtains the sample in a consecutive manner (i.e. on a first-come, first-served basis) after having established that the sample has fulfilled all the stipulated eligibility criteria and the recruitment process shall continue until the desired sample size has been achieved (Bowers et al., 2011; Bujang, 2017). In addition, consecutive sampling is also designed to recruit all the eligible subjects that can be ranked in a specified order, which is usually based on the earliest date and time. For example, only those patients who have fulfilled all eligibility criteria and are being notified for an earlier appointment at a hospital or in the clinic can be invited into the study within a stipulated timeframe (Chew et al., 2013; Bujang et al., 2015).

It is already well-known that the recommended probability sampling technique for an ordered population is systematic sampling (Fraenkal and Wallen, 2006). However, a major obstacle to adopting the systematic sampling technique in a clinical survey is the difficulty of prespecifying or setting the interval k since in some cases the total population number is not known. Therefore, the researcher may opt to apply a consecutive sampling technique even though it is a type of non-probability sampling. Therefore, this study aims to assess the viability of adopting consecutive sampling in lieu of systematic sampling by determining the accuracy of the sample estimates obtained by consecutive sampling. This can be achieved by comparing the sample statistics produced by consecutive sampling versus those produced by systematic sampling, and finally, a comparison is then made of all the sample estimates with the true value of the parameter in the target population.

Figure 1 is shown to visualize the recruitment procedure for both consecutive sampling and systematic sampling. Say, a population size (N) is designated as 8 and sample size (n) is designated as 4. For consecutive sampling, a researcher will recruit subject 1, subject 2, subject 3, and subject 4 to comply with the requirements on a ‘first-come and first-served basis. On the other hand, for systematic sampling, a researcher shall first and foremost have to calculate the value of interval (k), such as $N/n = 2$, and then shall perform the first

subject selection by random. Say, if the first chosen subject by random is 2, this means that the researcher shall recruit a series of subjects as subject 2, subject 4, subject 6, and subject 8.

Figure 1: Sample selection based on consecutive sampling and systematic sampling

As consecutive sampling is nonetheless still a type of non-probability sampling, it is therefore always necessary to determine to what extent it is able to generate accurate sample estimates for the target population. In such cases, it is, therefore, necessary to compare the sample statistics with the true population parameters which have been obtained between both consecutive sampling and systematic sampling techniques, through an initial simulation process and subsequently a validation process by using real-life data sets. Such findings are important for determining to what extent, within the remit of a sufficiently large pre-specified sample size, the estimates derived from the consecutive sampling are also close approximations of the statistics derived from the systematic sampling, and hence they shall be regarded as equally accurate and valid for representing the true values of the parameters of the target population as those of systematic sampling do. However, it should be emphasized that there is still a limitation on the generalizations about the population that can be made if consecutive sampling is used as this is a non-probability sample.

2. Study Methods

This study was conducted using two approaches. The first approach involved a simulation analysis and the second approach involved a validation technique based on a comparison of the estimates obtained by using real-life data sets.

2.1 First Approach: Simulation Analysis (This simulation analysis was conducted based on three distribution data sets; namely, normally distributed, exponentially distributed, and distributed as Poisson)

Step 1: Generate five sets of population data based on three different statistical distributions with a selected parameter as presented in Table 2. The total population size (N) is set at 3000.

Step 2: Design a sampling strategy by using systematic sampling initially and then by using consecutive sampling. The systematic sampling procedure for recruiting subjects is based on prespecifying the interval, k^{th} of five, and the first patient is selected at random from a single random number between one and k. For systematic sampling, the sample size is set at 600, and thus, therefore $k = 3000/600 = 5$. For consecutive sampling, the selection of subjects is performed in a consecutive manner (i.e on a first-come, first-served basis).

Step 3: Calculate the statistics for the mean and standard deviation for each sampling procedure. In each population, there were seven sub-samples as *n*, are 30, 50, 100, 200, 300, 500, and 1000.

Step 4: Iterate this procedure 1000 times. For each of the 1000 iterations, the mean square error (MSE) will be calculated for seven sub-samples within each population based on the two different sampling techniques. The comparisons will be made based on MSE from the two sampling techniques. Altogether, there are a total of 70 comparisons (five statistical distributions × two parameters × seven sub-samples).

2.2 Second Approach: Comparison of sample statistics and parameters between a sample obtained from consecutive sampling and another sample obtained from systematic sampling via a validation process

A validation process was conducted to determine the degree of proximity between population parameters and the sample estimates derived from samples obtained by using the two different sampling methods, namely: consecutive sampling and systematic sampling. This validation was performed in two different populations, by conducting four statistical analyses on six to eight samples. Both the definitions and explanations of the three key component data (namely: population data, statistical analysis, and sample size) are provided in Table 1. Data for this validation step have been obtained from "An Audit of Diabetes Control and Management (ADCM) 2009", which involved the collection of data from all patients with diabetes mellitus within all government health clinics in Malaysia, during the year 2009 at a national level (Ismail et al., 2009).

Table 1: Definition and explanation of the population, statistical analysis and sample size for the validation.

Component	Definition and explanation
Population	
P1 = All patients with diabetes mellitus notified in Health Clinic A, from January 1st until December 31st 2009 (N=1688)	
P2 = All patients with diabetes mellitus notified in Health Clinic B, from January 1st until December 31st 2009 (N=1986)	
Statistical analysis	
Descriptive	Mean (μ) of HbA1c for overall
Descriptive	Mean (μ) of HbA1c among male
Correlation (Spearman)	Correlation (*r*) between age and HbA1c
Multivariate (ANCOVA)	Marginal mean (μ) of HbA1c among male after controlled for age
Sample size	
Sample of 30, 50, 100, 150, 200, 300, 500 and 1000	

This study selected two tertiary health clinics with a relatively large number of patients (i.e. >1000 patients) so a series of analyses for the validation was conducted for a few sub-samples which displayed a wide range of different sizes from small to big. The consecutive sampling procedure for recruiting patients was based on a first-come, first-served basis, which depended on the notification date within the registry database. On the other hand, the systematic sampling procedure for recruiting patients was based on prespecifying the interval k^{th} of five with assumption $N = 1500$ and $n = 300$ ($N/n = 1500/300 = 5$). Then

randomly selecting the first patient who can belong to any rank number which is a single random number between one and *k*.

Four parameters were selected as a basis for comparison, namely: mean (μ) HbA1c, mean (μ) HbA1c among the male gender, correlation (*r*) between age and HbA1c, and adjusted mean calculated based on Analysis of Covariance (ANCOVA). The glycated haemoglobin A1c (HbA1c) test is a clinical measure that describes the patient's average level of blood sugar control over the past 2 to 3 months. A total number of 48 comparisons were performed (six sub-samples × four parameters × two populations). The sample estimates and reference parameters obtained from both sampling techniques were then compared.

The comparisons were made based on the selection of a sampling technique that can produce statistics with the minimum bias (between parameter and statistics) on average from the six sub-samples. The analysis involves descriptive analysis such as the calculation of mean, a univariate analysis via the correlation test, and a multivariate analysis by the Analysis of Covariance. All the statistical analyses for this study were performed using R software (R Core Team, 2014) and SPSS (IBM Corp. Released 2011. IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.)

3. Results

Both Table 2 and Table 3 illustrate the findings obtained from a simulation analysis. For a comparison of means, systematic sampling offers a smaller value of MSE in 16 out of 35 comparisons while consecutive sampling offers a smaller value of MSE in 19 out of 35 comparisons. For a comparison of standard deviations, systematic sampling has reported smaller values of MSE in 24 out of 35 comparisons while consecutive sampling has reported smaller values of MSE in 10 out of 35 comparisons. There are a total of 35 comparisons of MSE on SD. In one of the comparisons, the MSE for systematic sampling and consecutive sampling are the same. Overall, systematic sampling has provided a slightly greater advantage by reporting smaller MSEs in 40 out of 70 (or 57.1%) of the comparisons while consecutive sampling has reported smaller MSEs in only 29 out of 70 (or 41.4%) of the comparisons.

Table 2: The comparison of statistics between statistics derived from systematic sampling and consecutive sampling, results from a simulation analysis based on normal, exponential and Poisson distributions.

Parameter	Sample size	Systematic Sampling		Consecutive Sampling	
		Mean (95% CI)	SD (95% CI)	Mean (95% CI)	SD (95% CI)
Normal Distribution μ = 80 σ = 5	30	80.014 (78.233 - 81.746)	4.934 (3.636 - 6.165)	79.985 (78.215 - 81.647)	4.946 (3.656 - 6.302)
	50	80.002 (78.632 - 81.352)	4.978 (4.012 - 6.051)	80.001 (78.729 - 81.238)	4.955 (3.895 - 5.940)
	100	80.017 (79.073 - 81.021)	4.971 (4.249 - 5.649)	80.007 (79.051 - 80.933)	4.984 (4.301 - 5.717)
	200	80.023 (79.371 - 80.730)	4.990 (4.485 - 5.515)	80.005 (79.325 - 80.684)	5.001 (4.519 - 5.498)
	300	79.998 (79.426 - 80.542)	4.993 (4.595 - 5.400)	80.012 (79.489 - 80.539)	5.006 (4.617 - 5.420)
	500	79.993 (79.598 - 80.412)	4.996 (4.706 - 5.292)	80.003 (79.580 - 80.462)	5.014 (4.695 - 5.293)
	1000	79.997 (79.697 - 80.304)	5.006 (4.800 - 5.228)	80.006 (79.706 - 80.301)	5.005 (4.813 - 5.214)
Normal Distribution μ = 80 σ = 15	30	80.042 (74.698 - 85.237)	14.803 (10.909 - 18.496)	79.955 (74.645 - 84.941)	14.838 (10.968 - 18.905)
	50	80.006 (75.895 - 84.057)	14.933 (12.037 - 18.153)	80.004 (76.188 - 83.713)	14.864 (11.684 - 17.819)
	100	80.052 (77.218 - 83.064)	14.913 (12.748 - 16.948)	80.021 (77.153 - 82.798)	14.951 (12.902 - 17.150)
	200	80.070 (78.113 - 82.189)	14.969 (13.456 - 16.546)	80.015 (77.974 - 82.052)	15.003 (13.557 - 16.495)
	300	79.993 (78.279 - 81.627)	14.979 (13.784 - 16.200)	80.035 (78.467 - 81.616)	15.018 (13.852 - 16.259)
	500	79.979 (78.793 - 81.236)	14.988 (14.118 - 15.876)	80.008 (78.739 - 81.386)	15.041 (14.085 - 15.880)
	1000	79.991 (79.092 - 80.912)	15.018 (14.401 - 15.684)	80.018 (79.118 - 80.903)	15.014 (14.438 - 15.641)
Normal Distribution μ = 80 σ = 25	30	80.070 (71.163 - 88.728)	24.671 (18.181 - 30.826)	79.925 (71.074 - 88.236)	24.729 (18.281 - 31.508)
	50	80.010 (73.159 - 86.761)	24.888 (20.062 - 30.255)	80.007 (73.647 - 86.188)	24.773 (19.473 - 29.698)
	100	80.086 (75.363 - 85.107)	24.856 (21.247 - 28.247)	80.035 (75.255 - 84.664)	24.918 (21.504 - 28.583)
	200	80.116 (76.855 - 83.648)	24.948 (22.427 - 27.577)	80.025 (76.623 - 83.421)	25.005 (22.596 - 27.492)
	300	79.989 (77.131 - 82.711)	24.965 (22.973 - 27.000)	80.058 (77.445 - 82.694)	25.029 (23.086 - 27.098)
	500	79.965 (77.988 - 82.060)	24.980 (23.530 - 26.460)	80.013 (77.898 - 82.310)	25.068 (23.476 - 26.466)
	1000	79.984 (78.487 - 81.519)	25.030 (24.002 - 26.140)	80.030 (78.529 - 81.505)	25.023 (24.063 - 26.068)
Exponential Distribution λ = 0.0125	30	79.156 (54.180 - 110.300)	76.836 (45.789 - 124.763)	80.379 (53.278 - 109.967)	77.765 (45.269 - 117.010)
	50	80.012 (60.163 - 102.489)	78.554 (53.473 - 109.301)	79.939 (59.505 - 102.181)	78.297 (51.871 - 111.098)
	100	80.352 (66.199 - 95.804)	79.783 (59.519 - 102.411)	80.081 (65.625 - 95.991)	79.245 (58.140 - 103.558)
	200	80.158 (70.035 - 91.483)	79.690 (66.141 - 96.548)	80.243 (69.416 - 92.091)	79.700 (65.563 - 96.726)
	300	80.125 (71.531 - 88.944)	79.876 (68.756 - 93.617)	80.141 (71.617 - 89.699)	79.810 (67.705 - 93.785)
	500	80.113 (73.317 - 87.362)	79.987 (70.997 - 90.226)	80.063 (73.359 - 87.045)	79.859 (70.176 - 90.819)
	1000	79.904 (74.870 - 84.991)	79.844 (73.508 - 86.659)	79.971 (75.237 - 84.665)	79.988 (73.449 - 86.807)
Poisson Distribution λ = 4	30	4.026 (3.367 - 4.768)	1.991 (1.476 - 2.548)	3.992 (3.267 - 4.734)	1.972 (1.460 - 2.527)
	50	3.987 (3.480 - 4.561)	1.981 (1.600 - 2.404)	3.986 (3.440 - 4.540)	1.980 (1.596 - 2.419)
	100	3.999 (3.630 - 4.380)	1.996 (1.713 - 2.292)	3.993 (3.590 - 4.410)	1.989 (1.705 - 2.305)
	200	3.991 (3.725 - 4.310)	1.994 (1.791 - 2.198)	4.001 (3.715 - 4.280)	1.999 (1.791 - 2.222)
	300	4.000 (3.787 - 4.213)	1.998 (1.830 - 2.175)	4.002 (3.770 - 4.237)	2.001 (1.823 - 2.187)
	500	4.000 (3.826 - 4.170)	1.998 (1.873 - 2.132)	4.001 (3.826 - 4.180)	2.001 (1.867 - 2.136)
	1000	4.000 (3.890 - 4.117)	2.001 (1.909 - 2.095)	4.002 (3.876 - 4.123)	2.002 (1.910 - 2.096)

Table 3: Comparison of mean square error between systematic sampling and consecutive sampling based on normal, exponential and Poisson distributions on different sample sizes.

Parameter	Sample size	Systematic Sampling		Consecutive Sampling	
		MSE for Mean	MSE for SD	MSE for Mean	MSE for SD
Normal Distribution μ = 80 σ = 5	30	0.7911	**0.3995**	**0.7632**	0.4715
	50	0.4779	**0.2606**	**0.4483**	0.2612
	100	0.2492	0.1259	**0.2248**	**0.1234**
	200	0.1133	0.0621	**0.1115**	**0.0587**
	300	0.0751	0.0386	**0.0656**	**0.0373**
	500	**0.0391**	**0.0191**	0.0393	0.0200
	1000	0.0167	**0.0080**	**0.0155**	0.0081
Normal Distribution μ = 80 σ = 15	30	7.1197	**3.5951**	**6.8691**	4.2435
	50	**4.3012**	**2.3451**	4.0351	2.3509
	100	2.2427	1.1330	**2.0236**	**1.1110**
	200	1.0195	0.5588	**1.0034**	**0.5281**
	300	0.6755	**0.3475**	**0.5905**	0.3356
	500	**0.3520**	**0.1717**	0.3533	0.1798
	1000	0.1503	**0.0723**	**0.1397**	0.0726
Normal Distribution μ = 80 σ = 25	30	19.7770	**9.9863**	**19.0808**	11.7874
	50	11.9477	**6.5142**	**11.2085**	6.5304
	100	6.2297	3.1473	**5.6211**	**3.0861**
	200	2.8319	1.5521	**2.7871**	**1.4670**
	300	1.8764	0.9654	**1.6403**	**0.9323**
	500	**0.9776**	**0.4770**	0.9813	0.4993
	1000	0.4174	**0.2009**	**0.3880**	0.2016
Exponential Distribution λ = 0.0125	30	**203.8285**	380.6237	210.3496	**359.6566**
	50	**117.9314**	**226.2870**	121.7321	227.4735
	100	**58.3774**	**110.5930**	60.1264	121.8113
	200	**28.4590**	**57.6523**	30.2410	61.7133
	300	**18.3875**	**36.8262**	18.8156	39.6324
	500	**10.3341**	**21.8541**	10.5601	22.3768
	1000	**4.0119**	8.2338	4.1569	**7.9792**
Poisson Distribution λ = 4	30	0.1332	**0.0770**	**0.1326**	0.0777
	50	**0.0774**	**0.0430**	0.0803	0.0455
	100	**0.0371**	**0.0216**	0.0406	0.0225
	200	0.0201	**0.0104**	**0.0193**	0.0111
	300	**0.0119**	**0.0067**	0.0122	0.0073
	500	**0.0066**	**0.0036**	0.0068	0.0039
	1000	**0.0025**	0.0015	0.0028	0.0015

Note: MSE with bold refers as smaller MSE, a comparison between systematic sampling versus consecutive sampling
For means comparisons, systematic sampling is smaller in terms of MSE in 16 out of 35 comparisons while consecutive sampling is smaller in 25 out of 35 comparisons.
For standard deviations, systematic sampling is smaller in terms of MSE in 19 out of 35 comparisons while consecutive sampling is smaller in 10 out of 35 comparisons.
For overall comparisons, systematic sampling is smaller in terms of MSE in 35 out of 70 comparisons while consecutive sampling is smaller in 34 out of 70 comparisons. One comparison of MSE was identical.

Table 4 presents the results of the validation process which involves a comparison of the sample estimates of population parameters with the true population parameters, which have been obtained by both consecutive sampling and systematic sampling. These sample estimates of population parameters obtained by both sampling techniques were also more closely approximating to the true population parameters when the sample size was at least 300, which is at least 15.1% (i.e. [300/1986] × 100%) to 17.8% (i.e. [300/1688] × 100%) of the population size. Table 5 shows the comparisons that were made between the population parameters and the sample estimates, which were obtained by both consecutive sampling and systematic sampling.

Table 4: Comparison of statistics and parameter between results from consecutive sampling (CS) and systematic sampling (SS) from various types of analysis

Parameter to be measured	Population	Sample	Parameter	*n*=30	*n*=50	*n*=100	*n*=150	*n*=200	*n*=300	*n*=500	*n*=1000
Mean (μ) HbA1c	A	CS	8.23	7.67	7.49	7.38	7.58	7.54	7.58	7.70	7.99
		SS	8.23	7.60	7.67	7.40	7.67	7.96	8.10		
	B	CS	8.11	7.16	7.90	7.75	7.80	7.92	7.93	8.08	8.14
		SS	8.11	7.95	7.97	8.41	8.37	8.33	8.29		
Mean (μ) HbA1c among male	A	CS	8.22	8.22	7.83	7.68	7.62	7.59	7.65	7.77	7.93
		SS	8.22	7.70	7.71	7.68	8.05	7.99	8.12		
	B	CS	8.00	6.80	7.89	8.02	7.99	8.05	7.89	7.93	8.01
		SS	8.00	8.00	8.13	8.35	8.22	8.26	8.18		
Correlation (*r*)[a] between age and HbA1c	A	CS	-0.25	-0.22	-0.33	-0.28	-0.21	-0.21	-0.16	-0.22	-0.26
		SS	-0.25	-0.14	-0.15	-0.19	-0.19	-0.28	-0.25		
	B	CS	-0.22	-0.09	-0.08	-0.07	-0.14	-0.23	-0.25	-0.25	-0.20
		SS	-0.22	-0.02	-0.30	-0.38	-0.24	-0.23	-0.18		
ANCOVA (marginal mean, μ)	A	CS	8.24	8.22	7.87	7.72	7.64	7.61	7.67	7.80	7.98
		SS	8.24	8.15	7.82	7.66	8.00	7.98	8.11		
	B	CS	8.02	6.76	7.89	8.01	7.99	8.04	7.92	7.96	8.04
		SS	8.02	7.99	8.25	8.38	8.22	8.24	8.17		

Note: Statistics for systematic sampling were calculated until sample size of 300 due to limited sample size.
[a]Spearman's correlation test was applied instead of Pearson's correlation test since parametric assumption was violated.

Table 5: Comparison of differences between the estimates and parameter between sample derived from consecutive sampling (CS) and systematic sampling (SS)

Parameter to be measured	Population	Sample	*n*=30	*n*=50	*n*=100	*n*=150	*n*=200	*n*=300	Mean[a]
Mean (μ) HbA1c	A	CS	0.56	0.74	0.85	0.65	0.69	0.65	0.69
		SS	0.63	0.56	0.83	0.56	0.27	0.13	**0.50**
	B	CS	0.95	0.21	0.36	0.31	0.19	0.18	0.37
		SS	0.16	0.14	-0.3	-0.26	-0.22	-0.18	**0.21**
Mean (μ) HbA1c among male	A	CS	0	0.39	0.54	0.6	0.63	0.57	0.46
		SS	0.52	0.51	0.54	0.17	0.23	0.1	**0.35**
	B	CS	1.2	0.11	-0.02	0.01	-0.05	0.11	0.25
		SS	0.00	-0.13	-0.35	-0.22	-0.26	-0.18	**0.19**
Correlation (*r*) between age and HbA1c	A	CS	-0.03	0.08	0.03	-0.04	-0.04	-0.09	**0.05**
		SS	-0.11	-0.10	-0.06	-0.06	0.03	0.00	0.06
	B	CS	-0.13	-0.14	-0.15	-0.08	0.01	0.03	0.09
		SS	-0.20	0.08	0.16	0.02	0.01	-0.04	0.09
ANCOVA (marginal mean, μ)	A	CS	0.02	0.37	0.52	0.60	0.63	0.57	0.45
		SS	0.09	0.42	0.58	0.24	0.26	0.13	**0.29**
	B	CS	1.26	0.13	0.01	0.03	-0.02	0.10	0.26
		SS	0.03	-0.23	-0.36	-0.20	-0.22	-0.15	**0.20**

Note: [a] Reported absolute value of mean where values in bold refers to smaller value
Systematic sampling has smaller bias (difference between parameter and statistics in average) in six out of eight comparisons

Results obtained from these comparisons revealed that systematic sampling was able to provide more accurate estimates of the population parameters than consecutive sampling because it had shown six out of eight comparisons with the minimum differences and also one of the comparisons had ended up in a tie. These findings illustrated an important observation whereby systematic sampling is shown to yield smaller differences between the

sample estimates and the population parameters than consecutive sampling does, indicating that it can provide a higher level of precision in these sample estimates.

4. Discussions

Due to both cost and time constraints, consecutive sampling is often deployed for recruiting a sample in a survey, especially within the medical discipline (Bujang, 2017). However, as consecutive sampling is a form of non-probability sampling which can likely be introducing bias when collecting a sample; it is, therefore, necessary to assess and evaluate the accuracy of sample estimates which have been derived from consecutive sampling. Various studies were conducted to determine whether or not the different sampling methods can potentially affect the generalisability of these findings. A study conducted by Howes (1985) had previously stated that the sampling technique would not have much influence on the study outcome so long as the researchers were able to control those variables that could potentially introduce bias (Howes et al., 1985).

Results obtained from the simulation analysis in this study further substantiated this statement, which revealed that the study findings remain unaffected regardless of whether consecutive sampling or systematic sampling technique was deployed for sample recruitment, as long as the subjects have been randomly arranged in an ordered sequence (in other words, there is no pre-existing systematic pattern that appears in the order). On the other hand, various other research studies also found that the choice of sampling technique could potentially influence the accuracy of the sample estimates, in that the use of a probability sampling method (such as systematic sampling) can potentially improve the accuracy and precision of the sample estimates obtained (Yeager et al., 2011; Erens et al., 2014).

However, the results based on the simulation analysis of this study did not provide conclusive evidence that consecutive sampling is equivalent to systematic sampling. It is well-known that systematic sampling is proven to be an unbiased and efficient sampling technique. Consecutive sampling is regarded as a type of non-probability sampling and thus it will not be possible to defy its inherent bias. Hence, the ultimate aim of this paper is not to demonstrate equivalence between consecutive sampling and systematic sampling. Nonetheless, this paper hopes to illustrate by using real examples that although consecutive sampling is a form of non-probability sampling, the statistics derived from this sampling technique can still be likely to produce highly accurate statistics especially when its sample size is sufficiently large.

To this end, this paper has included an additional validation approach to determine to what extent the statistics derived from samples are close approximations to their respective parameters of the target population. In inferential statistics, the p-value is always regarded as an indicator that provides evidence for making inferences from raw observations. However, no one will know for sure whether the inference is valid or not unless a population or census study is conducted to validate the accuracy of these sample estimates. Again, this explains why this study also incorporates a validation process to validate these sample statistics apart from conducting an initial simulation analysis.

Since the true values of these population parameters are already known, the purpose of this validation process is to validate the sample statistics against the true population parameters. Hence, the purpose of validation, in this case, is not for making inferences, which means that the indicators such as the absolute differences between sample estimates and population parameters (i.e. sample statistics minus population parameters) were being calculated, instead of p-values and their 95% confidence intervals.

From the results, it can be deduced that even though consecutive sampling is a type of non-probability sampling technique, the sample estimates derived from this sampling technique can be close approximations to the population parameters especially if the sample size is adequately large. This is an important finding obtained from this study, which conducted a validation procedure by using four different types of statistical analysis, including both univariate and multivariate analyses. This finding was also found to be consistent with those of previous studies whereby the sample statistics derived from consecutive sampling were found to be almost equivalent to the true population parameters especially when the sample size was sufficiently large (Bujang et al., 2012; Bujang et al., 2015).

Based on the theory of probability sampling, the sample estimate derived from the systematic sampling method is both unbiased and efficient because it is a type of probability sampling (Taro, 1967). Therefore, based on the validation results of this study, our findings have shown that sample statistics derived from systematic sampling provided better estimates when compared to those obtained from consecutive sampling. Hence, by basing on real-life data sets, this study has successfully contended that apart from providing an unbiased estimate, a probability sampling technique such as systematic sampling can also produce better and more accurate estimates than a non-probability sampling technique such as consecutive sampling.

In conclusion, it is already well-known that sample statistics derived from the systematic sampling method will yield better and more accurate estimates than those from the consecutive sampling method. However, consecutive sampling will still be able to elicit a sample that can provide estimates for an ordered population that consists of randomly-arranged units, especially if a relatively large sample is used. This statement will hold provided there is no pre-existing consecutive pattern for the arrangement of population data. If necessary, a runs test can always be applied to ensure there is no consecutive pattern that pre-exists in a string of subjects before deploying the consecutive sampling technique to recruit subjects (Bujang and Sapri, 2018). However, it should be emphasized that there is limitation in the use of consecutive sample in making statistical inferences about the population.

One of the major limitations of this study is that the simulation was conducted only on a few selected distributions with specific population parameters. Future studies may explore the possibility of investigating other findings which are based on various other simulation techniques and models in order to compare findings derived from both sampling techniques. Secondly, the validation was conducted by using only two real-life data sets from the medical field. However, it is beyond the scope of this study to determine the applicability of this observation within all the various fields of study by performing an audit of the various real-

life data sets from among all the various fields. Despite the above, cumulative evidence has supported the contention that when the sample size is adequately large, it becomes likely for those estimates that were derived from the samples to more closely mimic the true parameters of the intended populations (Bujang et al., 2012; Bujang et al., 2015; Stockwell and Peterson, 2001; Hernandez et al., 2006).

Acknowledgement

The authors wish to thank the Director General of Health, Ministry of Health Malaysia for permission to publish this paper. A special thanks to Mr Hon Yoon Khee for his effort to proofread the article. We also would like to thank Madam Nur Khairul Bariyyah Mohd Hatta and Ms Lim Chien Joo for their help in processing the manuscript for publication.

Conflict of Interest Statement

All the authors declared we have no conflict of interest.

Funding Statement

This study receives no funding

Literature Cited

BOWERS, D., HOUSE, A. and OWENS, D., 2011, Carrying out a systematic search: Getting started in health research, United Kingdom: John Wiley and Sons Ltd.

BUJANG, M.A., 2017, Enhancement of systematic sampling for clinical survey: Systematic sampling with consecutive approach, Selangor: Institute of Graduate Studies, Universiti Teknologi Mara.

_____, GHANI, P.A., ZOLKEPALI, N.A., SELVARAJAH, S. and HANIFF, J., 2012, A comparison between convenience sampling versus systematic sampling in getting the true parameter in a population: Explore from a clinical database: The Audit Diabetes Control Management (ADCM) registry in 2009. In: ICSSBE 2012. Proceedings, 2012 International Conference on Statistics in Science, Business and Engineering Empowering Decision Making with Statistical Sciences, 10–12 September 2012, Kedah, Malaysia: 499-503. doi:10.1109/ICSSBE.2012.6396615.

_____, MUSA, R., LIU, W.J., CHEW T.F., LIM, C.T.S. and MORAD, Z., 2015, Depression, anxiety and stress among patients with dialysis and the association with quality of life, Asian J Psychiatr, 18: 49-52.

_____, SA'AT, N., JOYS, A.R. and ALI, M.M., 2015, An audit of the statistics and the comparison with the parameter in the population. In: AIP Conference Proceedings, Shah Alam, Malaysia, 1682(1): 050019. doi:10.1063/1.4932510.

_____and SAPRI, F.E., 2018, An application of the runs test to test for randomness of observations obtained from a clinical survey in an ordered population, Malays J Med Sci, 25(4): 146-151. doi:10.21315/mjms2018.25.4.15

CHEW, B.H., SHARIFF-GHAZALI, S., MASTURA, I., HANIFF, J. and BUJANG, M.A., 2013, Age ≥ 60 years was an independent risk factor for diabetes-related complications despite good control of cardiovascular risk factors in patients with type 2 diabetes mellitus, Exp Gerontol, 48(5):485–491. doi: 10.1016/j.exger.2013.02.017.

ERENS, B., BURKILL, S., COUPER, M.P., CONRAD, F., CLIFTON, S., and TANTON, C., 2014, Nonprobability Web Surveys to Measure Sexual Behaviors and Attitudes in the General Population: A Comparison with a Probability Sample Interview Survey. J Med Internet Res, 16:e276-1–276-14. doi:10.2196/jmir.3382

FLAMAING, J., DE BACKER, W., VAN LAETHEM, Y., HEIJMANS, S., and MIGNON, A., 2015, Pneumococcal lower respiratory tract infections in adults: an observational case-control study in primary care in Belgium. BMC Fam. Pract. 16. doi:10.1186/s12875-015-0282-1.

FRAENKAL, J.R. and WALLEN, N.E., 2006, How to design and evaluate research in education, New York: McGraw-Hill.

HERNANDEZ, P.A., GRAHAM, C.H., MASTER, L.L. and ALBERT, D.L., 2006, The effect of sample size and species characteristics on performance of different species distribution modelling methods. Ecography, 29(5): 773-785. doi:10.1111/j.0906-7590.2006.04700.x

HOWES, B.L., DACEY, J.W.H. and WAKEHAM, S.G., 1985, Effects of sampling technique on measurements of porewater constituents in salt marsh sediments, Limnology and Oceanograph, 30(1): 221-227. doi:10.4319/lo.1985.30.1.0221

ISMAIL, M., CHEW, B.H., LEE, P.Y., AI, T.C., SHARIFF, G.S., and HANIFF, J. 2011, Control and treatment profiles of 70,889 adult type 2 diabetes mellitus patients in Malaysia - A cross sectional survey in 2009, Int J Collab Res Intern Med Public Health, 3(1): 98-113.

JENSEN, L.F., HVIDBERG, L., PEDERSEN, A.F., and VEDSTED, P., 2015, Symptom attributions in patients with colorectal cancer. BMC Fam Pract, 16:1–10.

POLIT, D.F. and BECK, C.T., 2010, Essentials of Nursing Research: Appraising Evidence for Nursing Practice. Lippincott Williams & Wilkins. pp. 311–312.

STOCKWELL, D.R.B. and PETERSON, A.T., 2001, Effects of sample size on accuracy of species distribution models, Ecol. Model, 148(1): 1–13. doi:10.1016/S0304-3800(01)00388-X

TARO, Y., 1967, Elementary sampling theory, 1st ed., Englewood Cliffs (NJ): Prentice-Hall.

WIEGNER, L., HANGE, D., BJORKELUND, C., JR and AHLBORG, G., 2015, Prevalence of perceived stress and associations to symptoms of exhaustion, depression and anxiety in a working age population seeking primary care - an observational study. BMC Fam Pract **16**, 38 doi:10.1186/s12875-015-0252-7

YEAGER, D.S., KROSNICK, J.A., CHANG, L.C., JAVITZ, H.S., LEVENDUSKY, M.S. and SIMPSER, A., 2011, Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples. Public Opin Quart, 75: 709-747. doi:10.1093/poq/nfr020