i) This review of the sample design is based on the Vanuatu 2010 HIES data. The only variable considered in this review is per capita total household expenditure (variable of interest), as in addition to being one of the main indicators derived from the HIES, it is likely highly correlated with many other variables of interest (e.g. poverty). From the 2010 dataset, using this variable of interest, a list of relevant indicators were calculated, those indicators provide information on:
   - (a)the status of the household expenditure distribution within each province,
   - (b) The efficiency provided by the 2010 HIES sample design
   - (c) The accuracy of the estimates calculated from the 2010 HIES dataset (especially the per capita household expenditure, our variable or interest)

ii) The original dataset has been trimmed using the variable of interest, the lowest and the highest percentiles (the 1% households with the lowest and highest per capita total household expenditure) were removed from the analysis (outliers). The dataset ends up with 4,289 households (given 4,377 households were completed).

iii) The 2010 Vanuatu HIES sample was based on a stratified multi stages selection
   - Stratification: geographical provinces (by urban / rural locations)
   - First stage of selection: Enumerations areas (EAs) with probability of selection proportional to size
   - Second stage: households, with uniform probability of selection within the EAs

iv) The mean and standard deviation indicate the status of the variable of interest within each strata. The intracluster correlation (ρ)[1], and the design effect (DEFF)[2] highlight the efficiency of the sampling strategy, and the standard error/relative standard error (SE/RSE)[3] of the variable of interest show its accuracy.

v) The purpose of this analysis is to get some insights from the 2010 HIES sample design in order to improve the 2018 survey. There is no point to improve the sample size in strata where the sample is not efficient (the gain in accuracy will be minor compared to the related cost).

vi) The challenge in the 2018 Vanuatu baseline survey:

[1] The ρ is a measure of how similar households are to each other within a given enumeration area compared to the overall variability of the population as a whole. It is used to calculate the precision loss due to clustering. Like the standard deviation, the ρ is considered to be a true population parameter, and therefore transferable between designs.

[2] Design effect (DEFF) that measures the loss in accuracy the 2 stages selection generates (when we compare it from a simple random selection - SRS). If the design effect equals to 3 it means that the sample size has to be multiplied by 3 in order to obtain the same accuracy than the SRS. A DEFF close to 1 is suitable, a DEFF that equals 1 means that the design effect generated by the 2 stages selection does not impact the accuracy of the estimates (The SRS would have provided the same accuracy). The DEFF usually ranges between 1 and 3.

[3] Relative standard error of the estimates (RSE): the small the RSE is, the most accurate the estimate will be. Usually a RSE lower than 5% provides a high degree of accuracy

- Meet precision targets in each strata (provincial level) including Penama where Ambae island has been evacuated at the time of the sample design[4].
- Acceptable sample size (due to budget constraints)
- Following international recommendations (12 months of field operation)
- Enhance the monitoring and supervision of the field staff and simplify management of the logistics in the field

⇨ Optimize the variance/cost ratio of the survey design

vii) Table 1 presents the Vanuatu 2010 HIES survey specifications, efficiency and accuracy in each strata (for the variable of interest). It shows that some improvements can be done in Torba, and Shefa rural (where the RSE is higher than 5%), and it shows a high intraclass correlation in Malampa, Shefa rural and Tafea (that lead to a high design effect in those stratas). In Torba, the high design effect comes from the high number of households interviewed in each selected EA (on average 33 households per selected EA in this strata were interviewed)[5].

- Torba: the sample size is good, there is just a need to reduce the number of households to interview within each strata (and in order to keep a similar sample size the number of EAs to select in the province will be increased)
- Malampa: given the high intracluster correlation in this province, a higher number of EAs to select is required (with the same number of households per EA to interview).
- Shefa rural: keep the same number of households to interview within each EA, and increase the number of EA to select (this will lead to a higher sample size)
- Tafea: similar to Malampa province, the high intraclass correlation indicates that the number of EAs to select has to be increased (therefore the sample size as well).

The sample size has to be increased in Malampa, Shefa rural and Tafea, for the rest, the 2018 design will have to be similar as 2010 (in order to provide at least the same level of accuracy).

viii) The 2018 Vanuatu base line survey follows the international recommendations in terms of data collection schedule (12-month coverage) and considers a better management and supervision of the field staff. In this context, the field staff will work by team, given that:

- A team is made of 1 supervisor (team leader) and 2 or 3 interviewers
- Each interviewer will be responsible for 5 interview per round
- A round of survey is a 1 week period
- 1 EA is covered during 1 round, after the round completion, the team moves to the next EA for the next round.
- A team complete 32 rounds during the 12 month field operation period (roughly every 2 rounds/2 weeks) of work is followed by 1 round/1 week of rest).

ix) Table 3 presents a survey schedule starting February 2019 and ending February 2020. During this period of 32 working weeks (corresponding to 32 different selected EAs) the teams will be on the field (a 3 weeks period of rest during Christmas period).

[4] Ambae island has been evacuated in July 2018 due to volcano eruption. Vanuatu Government evacuated the entire population in the closest islands of Santo, Maewo and Pentecost. The Ambae population stayed in refugee camp the time they take a decision on re settlement but at that time decision was made to leave the island empty.

[5] Torba province is scattered and made of many remote islands that were excluded from the 2010 HIES selection. Only EAs on the main islands were kept (small number of EAs) hence a high number of households were interviewed in those EAs in order to achieve the desirable sample size. In order to increase the quality of estimates in this province the number of EAs excluded from the selection will have to be limited (cost implication).

x) The number of interviewer by team and number of team by province will determine the total sample size within each province. A team made of 3 interviewers can achieve 480 households over the period, while a team of 2 interviewers can achieve only 320 cases.

xi) The intraclass correlation is used to calculate the precision loss due to clustering. Like the standard deviation, the ρ is considered to be a true population parameter, and therefore transferable between designs. We have to accept the hypothesis that this correlation factor has not changed during the period 2010-2018, and therefore can be used to predict DEFF and RSE for the next survey given an adjusted design (based on the conclusions provided by the 2010 design). Table 2 predicts the design effect and sampling error of the variable of interest given the new sample design that is based on:

- the sample size within each strata
- the number of teams within each strata
- the number of interviewers per team

In order to allow more flexibility in the sample size, it is preferable to set up some teams of 3 interviewers, that can achieve 480 households, which represent a good sample size for Torba and Sanma urban and some teams of 2 interviewers that will achieve 320 households each (2 teams will be required in other provinces).

xii) The proposed design in Table 2 shows a total sample size of 4,640 households and a higher level of accuracy of the estimate of the variable of interest in all the stratas. Only Shefa rural shows a RSE higher than 5%, which will be still acceptable. The high intraclass correlation in Shefa rural impacts the variance of the estimates and lead to an increase the sample size or a decrease of the number of households to interview per EA which is logistically and financially not recommended.

Both stata codes “Vanuatu 2010 HIES – sample size computation.do” and “Vanuatu 2010 HIES.do ” shows the analysis of the 2010 HIES sample design.

In addition to this, the excel spreadsheet “Vanuatu NSD baseline survey sample – sample size and selection” presents all the details and specifications of the 2019/2020 survey design.

Table 1 – Vanuatu 2010 HIES sample design specifications

	2009 Census		2010 HIES									
Strata	N	tot Eas	EA selected	n	Av. hh per EA	Mean	St. deviation	St. error	Relative St. Error	DEFF	DEFT	ρ
11 - Torba	1766	26	13	437	33.6	13,477	9,551	769	5.7%	3.78	1.68	8.5%
21 - Sanma - urban	2552	46	29	480	16.6	18,857	13,363	821	4.4%	2.21	1.35	7.8%
22 - Sanma - rural	6661	116	61	590	9.7	16,926	11,973	629	3.7%	1.78	1.28	9.0%
32 - Penama	6620	119	60	604	10.1	17,299	12,910	674	3.9%	1.81	1.28	8.9%
42 - Malampa	7991	146	59	578	9.8	14,640	9,913	613	4.2%	2.37	1.49	15.6%
51 - Shefa - urban	9054	132	57	567	9.9	21,487	14,275	792	3.7%	1.85	1.32	9.6%
52 - Shefa - rural	6876	117	53	516	9.7	17,817	14,222	1,054	5.9%	3.07	1.68	23.7%
62 - Tafea	5853	125	54	517	9.6	13,913	10,062	631	4.5%	2.20	1.43	14.0%

Table 2: 2019/2020 baseline survey specifications

Strata	nb round	nb team	nb HHs per round/team	sample size	SE SRS	DEFF	DEFT	Standard error	RSE
11 - Torba	32	1	15	480	379	2.19	1.48	561.04	4.16%
21 - Sanma - urban	32	1	15	480	559	2.09	1.45	808.67	4.29%
22 - Sanma - rural	32	2	10	640	453	1.81	1.34	609.11	3.60%
32 - Penama	32	1	10	480	557	2.25	1.49	834.73	4.83%
42 - Malampa	32	2	10	640	378	2.40	1.55	584.86	4.00%
51 - Shefa - urban	32	2	10	640	548	1.86	1.36	746.73	3.48%
52 - Shefa - rural	32	2	10	640	542	3.13	1.77	958.92	5.38%
62 - Tafea	32	2	10	640	378	2.26	1.50	568.62	4.09%

Table 3: 2019/2020 baseline survey schedule

starting day	weekid	Torba	Sanma			Penama	Malampa		Shefa				Tafea	
		team 11	team 21	team 22	team 23	team 31	team 41	team 42	team 51	team 52	team 53	team 54	team 61	team 62
28/01/2019	week0	Survey training (P Vila)												
04/02/2019	week0													
11/02/2019	week0	Logistics matters												
18/02/2019	week1	R1 1040012	R1 2080012	R1 2040241	R1 2070132	R1 3070012	R1 4010031		R1 5170011	R1 5160073	R1 5130011	R1 5130711	R1 6040011	R1 6080171
25/02/2019	week2	R2 1040011	R2 2080021	R2 2010011	R2 2070031	R2 3070031	R2 4010071	R1 4080011	R2 5170031	R2 5160083	R2 5130031	R2 5130741	R2 6040012	R2 6070071
04/03/2019	week3						R3 4010081	R2 4080041						
11/03/2019	week4	R3 1070021	R3 2080031	R3 2010031	R3 2070042	R3 3070061	R4 4010101		R3 5170043	R18 5140042	R3 5130041	R3 5130761	R3 6040031	R3 6090011
18/03/2019	week5	R4 1070041	R4 2080041	R4 2010061	R4 2070081	R4 3070091		R3 4080022	R4 5170051	R19 5160093	R4 5130061	R4 5130801	R4 6040061	R4 6090041
25/03/2019	week6			R5 2030021			R5 4020011	R4 4080061						
01/04/2019	week7	R5 1040013	R5 2080071	R6 2030041	R5 2070091	R5 3070111	R6 4020021	R5 4080071	R5 5170071	R20 5160101	R5 5130081	R5 5130822	R5 6040081	R5 6090022
08/04/2019	week8	R6 1040021	R6 2080091	R7 2030061	R6 2070111	R6 3070141	R7 4020051		R6 5170111	R21 5160122			R6 6040101	R6 6080041
15/04/2019	week9	Easter Break												
22/04/2019	week10													
29/04/2019	week11	R7 1010051		R8 2030081			R8 4020071	R6 4070031	R7 5160011	R3 5070061	R6 5130091	R6 5130841	R7 6040131	
06/05/2019	week12	R8 1010041	R7 2080092	R9 2030111	R7 2060183	R7 3070171	R9 4020101	R7 4070061	R8 5160031	R4 5060021	R7 5130112	R7 5130861		R7 6080032
13/05/2019	week13		R8 2080142		R8 2060181					R5 5060061			R8 6010031	R8 6080131
20/05/2019	week14					R8 3050011	R10 4020262	R8 4070101	R9 5160051		R8 5130141	R8 5130871	R9 6020031	
27/05/2019	week15	R9 1030011	R9 2080143	R10 2040021	R9 2060141	R9 3050022	R11 4020291	R9 4070131	R10 5160071	R6 5050011	R9 5130161	R9 5130901	R10 6010011	R9 6080011
03/06/2019	week16	R10 1030012	R10 208016:	R11 2040031	R10 2060131	R10 3050041				R7 5050081			R11 6010022	R10 6080121
10/06/2019	week17	R11 1030021				R11 3050061	R12 4020251	R10 4090031	R11 5120041	R8 5050101	R10 5130164	R10 5130911	R12 6100011	
17/06/2019	week18	R12 1030041	R11 208017:	R12 2040061	R11 2060091	R12 3060021	R13 4020221	R11 4090072	R12 5120081		R11 5130171	R11 5130941	R13 6110011	R11 6080091
24/06/2019	week19	R13 1030051	R12 208018:	R13 2040071	R12 2060082	R13 3060031		R12 4090081					R14 6110033	R12 6070031
01/07/2019	week20						R14 4020191		R13 5120101	R22 5160131	R12 5130172	R12 5130961		
08/07/2019	week21	R14 1040022	R13 208024:	R14 2040101	R13 2060071	R14 3070191	R15 4020121	R13 4040021	R14 5120141	R23 5160135	R13 5130201	R13 5130981	R15 6040161	R13 6070041
15/07/2019	week22	R15 1040023	R14 208031:	R15 2040131	R14 2060061	R15 3070211	R16 4020141	R14 4040041						R14 6070061
22/07/2019	week23	Independence Break												
29/07/2019	week24													

starting day	weekid	Torba	Sanma			Penama	Malampa		Shefa				Tafea	
		team 11	team 21	team 22	team 23	team 31	team 41	team 42	team 51	team 52	team 53	team 54	team 61	team 62
05/08/2019	week25	R16 1050041	R15 208034	R16 2040151	R15 2060051	R16 3080041	R17 4010011			R24 5160151	R14 5130221	R14 5131001		R15 6070021
12/08/2019	week26	R17 1050052		R17 2040181	R16 2060021	R17 3080051	R18 4010171	R15 4030221	R15 5120172	R25 5170121	R15 5130232	R15 5131011	R16 6050101	R16 6070171
19/08/2019	week27	R18 1050061	R16 208036						R16 5120181	R26 5170132			R17 6050121	
26/08/2019	week28	R19 1050071	R17 208038		R17 2060011		R19 4010151	R16 4050011			R16 5130261	R16 5131031		
02/09/2019	week29	R20 1050072		R18 2040211	R18 2050191	R18 3080091	R20 4010131	R17 4050031	R17 5150022		R17 5130292	R17 5131051		R17 6080281
09/09/2019	week30	R21 1050081	R18 208038	R19 2040222		R19 3080122		R18 4050041	R18 5150041	R9 5010021			R18 6050141	R18 6080291
16/09/2019	week31	R22 1050091	R19 208041		R19 2050171		R21 4030191	R19 4050062	R19 5150061	R10 5010041	R18 5130311	R18 5131073	R19 6050171	R19 6080061
23/09/2019	week32	R23 1050092		R20 2040252	R20 2050131	R20 3090051	R22 4030251	R20 4050081			R19 5130341	R19 5131081		
30/09/2019	week33		R20 208044	R21 2040261		R21 3090071	R23 4030261	R21 4050101	R20 5100011	R11 5010061	R20 5130361	R20 5131101	R20 6050191	R20 6080201
07/10/2019	week34	R24 1020012	R21 208045		R21 2050101			R22 4050121	R21 5100021	R12 5020012			R21 6050251	R21 6080242
14/10/2019	week35	R25 1020011		R22 2040291	R22 2050091	R22 3090111	R24 4030041		R22 5100071		R21 5130381	R21 5131131		
21/10/2019	week36		R22 208048			R23 3090141	R25 4030011			R13 5020031	R22 5130421	R22 5131141	R22 6050271	R22 6080261
28/10/2019	week37	R26 1040051	R23 208049	R23 2100091	R23 2050051		R26 4030061	R23 4060082	R23 5100101	R14 5030011	R23 5130431	R23 5131161	R23 6050291	R23 6060171
04/11/2019	week38	R27 1040052		R24 2100011	R24 2050031	R24 3090171		R24 4060041	R24 5100111	R15 5030041				
11/11/2019	week39		R24 208049	R25 2100061		R25 3090211	R27 4030071	R25 4060031		R16 5030061	R24 5130491	R24 5131191	R24 6050311	R24 6060151
18/11/2019	week40	R28 1050011	R25 208046	R26 2100082	R25 2020131		R28 4030101	R26 4060141	R25 5110021	R17 5040011	R25 5130521	R25 5131211	R25 6050321	R25 6060031
25/11/2019	week41	R29 1050012			R26 2020111	R26 3100012	R29 4030131	R27 4060161	R26 5110031					
02/12/2019	week42	R30 1050021	R26 208043	R27 2090021	R27 2020101	R27 3100031		R28 4060201	R27 5110061	R27 5170163	R26 5130541	R26 5131221	R26 6050031	R26 6060181
09/12/2019	week43		R27 208032	R28 2090041	R28 2020081			R29 4060211	R28 5110091	R28 5180011	R27 5130561	R27 5131242	R27 6050071	R27 6060131
16/12/2019	week44	Christmas break												
23/12/2019	week45													
30/12/2019	week46													
06/01/2020	week47		R28 208029		R29 2020061	R28 3100061					R28 5130591	R28 5131261	R28 6050211	R28 6060101
13/01/2020	week48	R31 1040023		R29 2090072		R29 3100072	R30 4040071	R30 4100011	R29 5090011	R29 5180021	R29 5130611	R29 5131271	R29 6050231	R29 6060111
20/01/2020	week49	R32 1040024	R29 208023	R30 2090081	R30 2020031		R31 4040091	R31 4100051		R30 5180062				
27/01/2020	week50		R30 208013		R31 2020011	R30 3100091	R32 4040111	R32 4100071	R30 5080011		R30 5130631	R30 5131283	R30 6070091	R30 6060082
03/02/2020	week51		R31 208011	R31 207010		R31 3100102			R31 5080031	R31 5180081	R31 5130661	R31 5131292	R31 6070141	R31 6060051
10/02/2020	week52		R32 208020	R32 207017	R32 2030141	R32 3100131			R32 5080051	R32 5180091	R32 5130691	R32 5131311	R32 6070151	R32 6060061