The Oncologist
The official journal of the Society for Translational Oncology

Supplemental Appendix for:
Lack of Understanding of Cervical Cancer and Screening Is the Leading Barrier to Screening Uptake in Women at Mid-Life in Bangladesh: Population Based Cross Sectional Survey
Susan Davis et al.

Sampling Weight for Bangladesh Mid-life Women's Health Study (BMWHS)

INTRODUCTION

The Bangladesh Mid-life Women's Health Study (BMWHS) was designed to be a nationally representative household survey of women aged 30-59 years of age. The main objectives of this survey were to investigate women's knowledge of cervical cancer (CCa) and breast cancer (BCa) screening; to understand the socio-economic and demographic issues which might act as barriers to early diagnosis of CCa and BCa; to document the prevalence of urinary incontinence (UI) and fecal incontinence (FI), and the risk factors for these conditions; and to determine the prevalence and severity of symptoms of the menopause and achieve a best estimate of the average age of menopause, taking into account the limitation of rural women being uncertain of their age. The survey design requirements for this study have been developed so that precise estimates can be generated for Bangladesh as a whole. The target population for this survey includes all women (3.24 million) [1], aged 30 to 59 years of old residing in 7 of the 32 districts where cervical cancer screening has been introduced.

STUDY DESIGN AND SAMPLING PROCEDURE

Between September 2013 and March 2014, we conducted the BMWHS, a population-based cross-sectional survey, utilizing a multistage cluster sampling technique (Figure 1). Bangladesh is divided into seven major administrative divisions which are made up of 64 administrative districts. Each division is divided into several districts and sub-districts. Within sub-districts, there are enumeration areas (EAs) both in urban and rural areas that are the smallest units with a defined area designated by the "Monitoring the Situation of Vital Statistics of Bangladesh" and provided by the Bangladesh Bureau of Statistics (BBS) [2]. In the BMWHS, district, EA and household were deemed as primary sampling unit (PSU) secondary sampling unit (SSU), and final sampling unit (FSU) respectively.

Of 64 districts, 32 had offered an opportunistic CCa screening program. As our primary aim was to determine the barriers to uptake of CCa screening, our sampling was restricted to these 32 districts. We randomly selected one district that had offered CCa screening from each division. The included districts (PSUs) were Barisal, Tangail, Comilla, Sathkhira, Rajshahi, Rangpur and Habigonj. The number of women recruited in each of the 7 selected districts was determined by the distribution of the 3.24 million women aged 30 to 59 years in those districts based on 2011 population census[1]. We also took into account the ratio of urban-rural women in the target age group within each district. The selected districts contained between 18 and 72 EAs (SSUs). Each EA includes, on average, 120 households (FSUs) [2]. As shown in Figure 2, when the total sample size of 1586 was distributed across the rural and urban areas of the 7 districts, the smallest number of women required to be recruited in any single area was 36. Thus, the household selection was an unequal probability systematic selection with 36 households per EA. Only one eligible woman was recruited from each household. If an index woman was not found in a selected household, the adjacent household was approached.

SAMPLE SIZE

The sample size for the BMWHS was calculated based on the estimated prevalence of CCa screening uptake among women at midlife. A sample size of 755 allowed us to estimate the prevalence of CCa screening uptake to within ±2% of the estimate of 8.6%. As the sample was selected using a cluster sampling, the sample size was multiplied by the design effect (deff), which commonly ranges from 1.5 to 3.0. In this study, deff was given a value of 2, as the population characteristics were almost homogenous when urban and rural EAs were considered independently [3]. The sample size was further increased by 5% to allow for contingencies. Hence, the final sample size was 1590 women.

Table 1 shows the divisions, selected PSUs, total number of SSUs, population and sampled SSUs.
Table 1. The distribution of sampled SSU and population by selected PSU

Divisions	Selected Districts (PSUs)	Total EAs (SSUs)	Population Size	Sampled EAs (SSUs)	Sample Size
Barisal	Barisal	72	360211	6	176
Chittagong	Comilla	36	757687	10	371
Dhaka	Tangail	28	610995	11	299
Khulna	Sathkhira	18	329803	6	162
Rajshahi	Rajshahi	66	433578	7	212
Rangpur	Rangpur	63	459880	7	225
Sylhet	Habigonj	28	286588	4	140
Total		**311**	**3238742**	**51**	**1586**

SAMPLING PROBABILITIES AND SAMPLING WEIGHTS

Due to unequal probabilities, sampling weights were required to ensure the actual representativeness of the sample at divisional level, district (PSU) level as well as EA (SSU) level.

The weighting process for BMWHS involved a three-step process: firstly, the design weight or base weight, calculated from all steps of random selection in the sample design; secondly, an adjustment for non-response by sampled households/individuals eligible for the survey; and finally, a post-stratification adjustment (calibration) of sample totals to the known population totals.

Design weight

The inverse of the unconditional probability of selection was the final selection weight (design weight) for each respondent, which is the product of the probabilities of selection associated with each stage of the design. In order to calculate the sampling weights, sampling probabilities were calculated separately for each sampling stage.

The BMWHS describes the following relevant components of the overall selection probability when the study follows a multi-stage sampling approach. The subscripts *a* and *k* (jointly for the *a-th* PSU and *k-th* SSU, respectively) in this description jointly correspond to the SSU, which is chosen in two sampling stages in selecting a respondent *R:*

$P^{(1)}_{ak}$ =Unconditional probability of selecting *a-th* PSU (geo-political area /cluster/district in which *R* lives) and *k-th* SSU (segment/EA in which *R* lives);
$P^{(2)}_{ak}$ =Conditional probability (given PSU and SSU selections) of selecting the household in which *R* lives.

Note that each of the selection events corresponding to these probabilities must occur in order for *R* to be selected in the BMWHS sample. Then the unconditional joint probability of selecting *R* (the *ak*-th person) into the BMWHS sample is

$$P_{ak} = P^{(1)}_{ak} \times P^{(2)}_{ak} \qquad \text{(i)}$$

Since the probability of joint sequential events is the unconditional probability of the first event in the sequence times the conditional probabilities of each subsequent event given the outcome of the prior events in the sequence, the associated design weight for *R* is:

$$D_{ak} = 1/P_{ak} = 1/ (P^{(1)}_{ak} \times P^{(2)}_{ak}) \qquad \text{(ii)}$$

As seen in Eq. (ii), we must determine each of the components of unconditional joint probability of selection *R* in order to compute *R's* design weight.

As an example, one SSU (SSU ID: 758) the Barisal division had two screened districts with total of 9702 HHs. However, one district that was randomly selected in our study had total of 72 SSUs with 7417 HHs. We used without-replacement simple random sampling (SRS) to select PSUs within first stage sampling strata. Then N_i = 7417 is the size measure (in number of households as of the last census 2011) for *R's* PSUs, I = 1 is the number of PSU chosen in the sampling stratum from which *R's* PSU was chosen, and the sum of the size measures for

all PSUs in that stratum is $\Sigma_a N_a = 9702$, the unconditional selection probability for *R's* will be, (all calculations have been presented in three decimal points).

$P^1_a = IxN_a/\Sigma_a N_a = 1x7417/9702 = 0.764$

If *R's* SSU is one of $K_a = 6$ SSUs chosen by (without-replacement) SRS from $S_a = 72$ SSUs in *R's* PSU, then the conditional probability (given PSU selection) of selecting *R's* SSU is,

$P^1_{k(a)} = K_a/S_a = 6/72 = 1/12 = 0.083$

and the unconditional joint probability of selecting *R's* PSU and R's SSU is,

$P^1_{ak} = P^1_a \times P^1_{k(a)} = (IxN_a/\Sigma_a N_a) \times (K_a/S_a)$

$= 0.764X\ 0.083 = 0.064$

As we used SRS to select households, the probability is the ratio of household sample size in each segment to the total number of households on the frame list for each segment. In our study $H_{ak} = 36$ households were selected by without-replacement SRS from $L_{ak} = 120$ listed households in the *ak-th* SSU in which *R* household is located, then the conditional probability of selecting *R's* household is,

$P^{(2)}_{ak} = H_{ak}/L_{ak} = 36/120 = 0.300$

Combining all of the probabilities for the selection process that led to choosing *R,* we have in summary then that *R's* unconditional overall selection probability is,

$P_{ak} = P^{(1)}_{ak} \times P^{(2)}_{ak} = 0.064x\ 0.300 = 0.019$

And the respondent design weight is,

$D_{ak} = 1/P_{ak} = 1/0.019 = 52.323$ (Integrated design weight for a respondent in a SSU in Barisal)

And the respondent normalized weight is,

$Y = D_{ak}*(n/N)$

Where $D_{ak} = 52.323$ (approx.)

$n = 1590$, and

$N = 68235$ (total number of women represented by the sample after scale weighting/integrated design weight)

$Y = 1.219$ (approx.)

Following table 2 shows the integrated design weight for BMWHS HHs/respondent:

Table 2. Integrated Design weight for BMWHS HHs/respondent

Randomly selected PSU/District	**Randomly selected SSU ID**	***HHs/Respondent's Design Weight (D_{ak})**	**Normalize Weight (Y)=[D_{ak}*(n/N)]**
Barisal	758	52.323	1.219
(6 SSUs)	763	45.347	1.057
	1388	37.062	0.864
	1393	63.224	1.473
	5	65.404	1.524
	510	60.1714	1.402
Tangail	189	36.937	0.861
(10 SSUs)	690	49.802	1.160
	691	33.616	0.783
	942	52.292	1.218
	1072	37.352	0.870
	1195	33.616	0.783
	1324	38.182	0.889
	314	49.802	1.160
	439	58.517	1.363
	818	57.272	1.334
Comilla	401	31.878	0.742

(11 SSUs)	403	31.347	0.730
	781	22.846	0.532
	1034	11.423	0.266
	152	30.815	0.718
	275	37.988	0.885
	531	20.189	0.470
	654	25.503	0.594
	780	26.565	0.619
	1410	37.191	0.867
	1411	34.003	0.792
Satkhira	329	16.972	0.395
(6 SSUs)	586	24.334	0.567
	706	15.745	0.367
	961	23.311	0.543
	330	24.538	0.571
	707	10.633	0.248
Rajshahi	92	87.008	2.027
(7 SSUs)	343	97.579	2.273
	598	85.381	1.989
	847	90.260	2.103
	972	81.316	1.894
	346	97.579	2.274
	470	49.602	1.156
Rangpur	238	50.120	1.168
(7 SSUs)	618	36.755	0.856
	868	60.144	1.401
	993	47.196	1.099
	1115	42.184	0.983
	738	31.325	0.729
	989	63.068	1.469
Habigonj	114	10.499	0.244
(4 SSUs)	871	27.805	0.647
	1248	21.972	0.511
	870	19.639	0.458

*Corresponding weight of a SSU is applicable for all 36 respondents of that particular SSU.

The adjustment for non-response

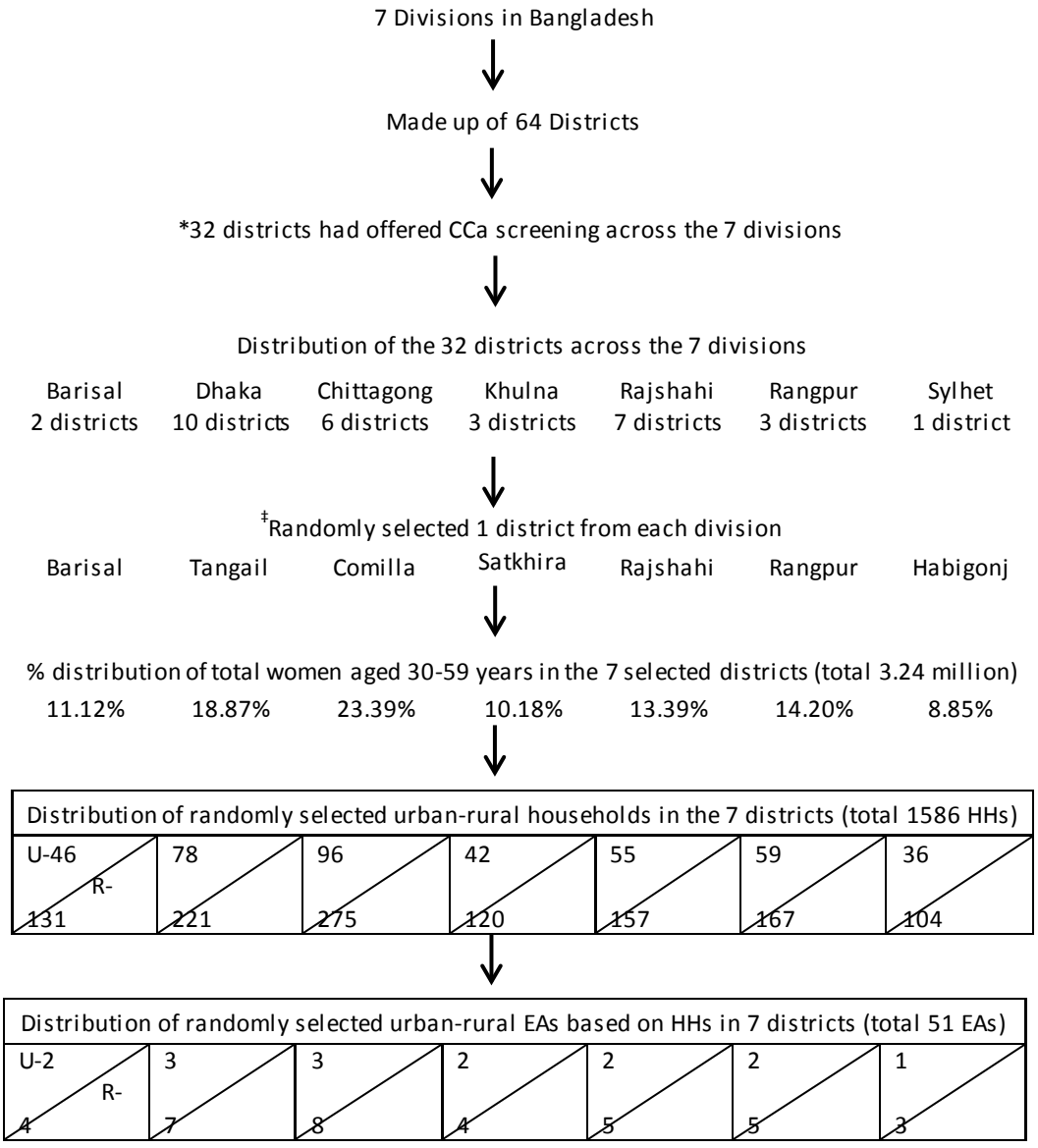
In the BMWHS, some households provided no data at all – *unit non-response* – while other households provided partial data – *item non-response*. There are three basic procedures to compensate unit non-response: i) non-response adjustment of the weights; ii) drawing a larger sample than needed and creating a reserve sample from which replacements are selected in case of non-response; and iii) substitution, the process of replacing a non-responding household with another household that was not sampled which is in close proximity to the non-responding household with respect to the characteristic of interest[4]. Substitution is a frequently used procedure in many surveys in developing countries and has been used in the BMWHS. Secondly, the standard method of compensating for item non-response is *imputation*, which is not applied in this study due to very low item non-response [4].

Post-stratification (calibration) adjustment

In principle, the goal of a calibration weight adjustment is to bring weighted sums of the sample data in line with the corresponding counts in the target population [5]. This procedure concurrently compensates for non-response and non-coverage. It adjusts the weighted sampling distribution for certain variables so as to conform to a known population distribution [4]. In the BMWHS, age and education were considered as the significant variables to adjust for. Known population distribution for age and education were taken from last

population census 2011**.** However, we were not able to calibrate design weight due to lack of age-sex specific education census data from each study PSU and SSU.

Figure 1. Sampling procedure of the BMWHS



Note:
*List of 32 districts were collected from Directorate General of Family Planning (DGFP), Ministry of Health& Family Welfare, Bangladesh;
‡The single district from Sylhet was selected.
U = Urban
R = Rural

(**Source:** Adapted from Islam et al. Bangladesh midlife women's health study (BMWHS): methods, challenges and experiences. Maturitas 2015; **80:** 89–94)

REFERENCES

1. Bangladesh Bureau of Statistics. Population and housing census 2011: Socio-economic and demographic report. National series, volume-4. Bangladesh bureau of statistics (BBS) statistics and informatics division (SID) ministry of planning government of the people's republic of

Bangladesh. http://203.112.218.66/webtestapplication/userfiles/image/bbs/socio_economic.Pdf. Accessed December 10, 2014.

2. Bangladesh Bureau of Statistics. Monitoring the situation of vital statistics of Bangladesh (MSVSB) project. Bangladesh bureau of statistics, statistics division, ministry of planning, government of the people's republic of Bangladesh, dhaka, Bangladesh. 2011

3. Henry GT, Sage Publications inc. Practical sampling. Applied social research methods series v 21. 1990.

4. Division UNS, Programme NHSC. Household surveys in developing and transition countries. United Nations Publications; 2005.

5. Global Adult Tobacco Survey Collaborative Group. Global adult tobacco survey (gats): Sample weights manual, version 2.0. Atlanta, GA: Centers for Disease Control and Prevention. 2010. http://www.who.int/tobacco/surveillance/en_tfi_gats_sampleweightsmanual_v2_final_15dec2010.pdf Accessed December 28, 2012.