BS 728
Homework 2
Irene Hsueh

Question 1

PSU = primary sampling unit
SSU = secondary sampling units
TSU = tertiary sampling units
PPS = proportional to population size
CEB = census enumeration block

Rural Area Sampling – Stratified Sampling
1) creating strata of villages with probability PPS
2) systematic sampling in each strata

Urban Area Sampling – Cluster Sampling
1) PSU: PPS sampling of wards
2) SSU: one CEB per ward
3) TSU: households in each CEB

Urban Area 8 Cities – Cluster Sampling
1) PSU: slum and non-slum CEB
2) SSU: households in each CEB

Part A
        Different sampling strategies were used for urban and rural areas because the 2001 census information for urban areas was not published at the time the DHS was conducted so listing all the households in each large urban area proved extremely difficult. Stratified sampling was used for rural areas, as each strata was systematic sampled. Cluster sampling was used for urban areas because at each stage, a simple random sampling was done on the primary and secondary (and tertiary for the eight cities) units. Stratified and clustering sampling are different because in stratified sampling, every stratus is sampled but in clustering sampling, random sampling will determine which clusters are selected.

Part B
The sampling frame of the rural areas was the people recorded in the 2001 Census list of Villages.

Part C
Stratification was used in the rural sample to make sure villages with different socioeconomic statuses were represented in the sample. The first level of stratification was geographic.

Question 3

Part A
The overall estimate of TB prevalence is 2.20%.

Part B & C

	Overall (N=50000)	TB (N=1,101)	No TB (N=48,899)	p-value
Male (N, %)	25,253 (50.51%)	537 (48.77)	24,716 (50.55)	0.2451
Female (N, %)	24,747 (49.49%)	564 (51.23)	24,183 (49.45)	
Age (mean, 95% CI)	26.80 (26.63, 26.97)	26.03 (24.87, 27.17)	26.82 (26.65, 26.99)	0.1784
Age Category				0.3283
0-4	5,246 (10.49%)	116 (10.54%)	5,130 (10.49%)	
5-14	11,225 (22.45%)	272 (24.70%)	10,953 (22.40%)	
15-39	20,849 (41.70%)	442 (40.15%)	20,407 (25.37%)	
$\geq$40	12,675 (25.4%)	271 (24.61%)	12,404 (25.37%)	
Missing	5			
BMI (mean, 95% CI)	21.07 (20.99, 21.15)	20.23 (19.51, 20.94)	21.09 (21.01, 21.17)	0.0183
BMI category				0.0014
Malnourished	1,410 (12.40%)	49 (18.28%)	1,361 (12.25%)	
Underweight	1,916 (16.84%)	49 (18.28%)	1,867 (16.81%)	
Healthy	6,316 (55.53%)	147 (54.85%)	6,169 (55.54%)	
Overweight	1,317 (11.58%)	21 (7.84%)	1,296 (11.67%)	
Obese	416 (3.66%)	2 (0.75%)	414 (3.73%)	
Missing	38,625			
Windows in home (N, %)	37,039 (74.12%)	681 (61.85%)	36,358 (74.40%)	<0.0001
Missing	30			
Cook with wood fuel (N, %)	22690 (45.38%)	588 (53.41%)	22,102 (45.20%)	<0.0001

Table 1a. Analysis not accounting for sampling design.

Part D

	OR	95% CI	p-value
Female vs Male	1.073	0.952, 1.209	0.2496
Age Category			
0-4	0.973	0.781, 1.213	0.8808
5-14	1.074	0.905, 1.274	0.8678
15-39	0.987	0.847, 1.150	0.4118
$\geq$40	REF		
Windows	0.591	0.520, 0.672	<0.0001
Wood Fuel	1.230	1.087, 1.392	0.0011

Table 2a. Regression not accounting for study design

Question 4

Part A

The overall estimate of TB prevalence is 2.31%.

Part B & C

	Overall (N= 512333)	TB (N= 11845)	No TB (N= 500488)	p-value
Male (N, %)	256480 (50.06%)	5596 (47.24%)	250884 (50.12)	0.1399
Female (N, %)	255853 (49.94%)	6249 (52.76%)	249604 (49.87)	
Age (mean, 95% CI)	26.18 (25.94, 26.43)	25.47 (23.97, 26.97)	26.21 (25.96, 26.45)	0.3450
Age group				0.5800
0-4	58365 (11.39)	1515 (12.79)	56850 (11.36)	
5-14	121878 (23.79)	2893 (24.42)	118985 (23.78)	
15-39	205376 (40.09)	4485 (37.86)	200892 (40.4)	
$\geq$40	126663 (24.73)	2952 (24.92)	123711 (24.72)	
Missing	5			
BMI (mean, 95% CI)	20.52 (20.42, 20.63)	19.54 (18.94, 20.15)	20.54 (20.44, 20.65)	0.0014
BMI category				0.0152
Malnourished	17296 (15.07)	647.28 (22.63%)	16649 (14.89%)	
Underweight	22560 (19.66)	661.19 (23.12%)	21899 (19.57%)	
Healthy	60822 (53.00)	1359 (47.52%)	59463 (53.14)	
Overweight	10763 (9.37)	168.95 (5.91%)	10594 (9.47%)	
Obese	3320 (2.89)	23.58 (0.82%)	3296 (2.95%)	
Missing	38625			
Windows in home (N, %)	343541 (67.09)	6330 (53.44)	337210 (67.41)	<0.0001
Missing	30			
Cook with wood fuel (N, %)	253289 (49.44)	6039 (50.98)	247250 (49.40)	0.5591

Table 1b. Analysis accounting for sample design.

Part D

	OR	95% CI	p-value
Female vs Male	1.122	0.963, 1.307	0.1399
Age			
0-4	1.062	0.801, 1.407	0.6767
5-14	0.972	0.776, 1.219	0.8085
15-39	0.933	0.765, 1.138	0.4938
$\geq$40	REF		
Windows	0.552	0.45, 0.68	<0.001
Wood Fuel	0.947	0.758, 1.182	0.6299

Table 2b. Regression accounting for study design

Question 5

Having windows in the home is associated with TB. Using survey weights, individuals with TB only made up 2.31% of the total sample, so there may not be enough power to detect potential associations with TB.