

Chapter 4. Within Imputation processes

First, missing data on demographic, socio-economic and self-reported medical conditions in MCBS and NHANES are imputed using sequential regression multiple–imputation procedure. Covariates listed in Table 1 and self-reported medical conditions were missing for some subjects in both NHANES and MCBS data files. A sequential regression multiple imputation procedure (Kennickel (1992), Van Buuren and Oudshoorn (2000) and Raghunathan et al (2001)) as implemented in the software package IVEWARE (Raghunathan et al (2002)) was used multiply impute the missing values. These are iterative procedures in which the missing values in each variable are imputed conditional on all other variables using appropriate regression models. Random draws from an approximate predictive distribution of the missing values under these models are then used as imputations.

Surveys for community dwellers and institutionalized populations differ, hence the sets of common MCBS-NHANES variables were different for institutionalized and community populations. We separately imputed missing values of common covariates in NHANES, MCBS-community, and MCBS-institutionalized populations. The rates of missing values for varied by survey ranging between [0%-81%] and were generally highest in the MCBS institutionalized population. Table shows percent of missing values in each survey for year missing values for year 2009. Some variables with the highest rate of missingness in MCBS were imputed by combining the community and institutionalized MCBS populations to increase stability.

Table 1a. Rates of missingness across the surveys (2009)

Variable name	Label	NHANES	MCBS	MCBS Inst
age	Age (continuous)	[0.0%]	[0.0%]	[0.0%]
asthma_emphysema	Asthma/Emphysema/COPD (combined in MCBS)		[0.0%]	
bc_taken	Last blood cholesterol check	[2.0%]	[3.0%]	
bp_taken	Last blood pressure check	[0.0%]	[0.7%]	
comphealth	Health compared to 1 year ago	[0.0%]	[0.7%]	[4.3%]
cost	Expenditure label		[0.0%]	[0.0%]
didserv	Served in Armed Forces	[0.0%]	[0.2%]	[7.0%]
dif_lift	Difficulty lifting/carrying 10 pounds		[0.7%]	[2.0%]
dif_rch	Difficulty reaching up over head	[0.1%]		

Variable name	Label	NHANES	MCBS	MCBS Inst
dif_stoop	Difficulty stooping/crouching/kneeling	[0.0%]	[0.7%]	[1.6%]
dif_walk	Difficulty walking 1/4 mi	[0.5%]	[0.8%]	[1.7%]
ed5	Education	[0.3%]	[0.6%]	[12.3%]
eversmoke	Ever smoked		[0.0%]	[21.8%]
flushot	Flu shot last year		[1.1%]	[10.1%]
hasjob	Has job	[0.0%]	[0.3%]	
havecare	Have particular place for medical care	[0.0%]	[0.7%]	
healthstat	General health status	[0.0%]	[0.8%]	[1.6%]
hearing	Quality of hearing	[0.2%]	[0.5%]	
hearingaid	Wear hearing aid	[32.1%]	[0.8%]	[3.6%]
height	Height (cm)	[2.5%]	[0.2%]	[0.0%]
hyst	Had Hysterectomy	[6.8%]	[8.8%]	[17.5%]
i_days	Number of Days Patient in Institution	[0.0%]	[0.0%]	[0.0%]
inpatnights	Inpatient nights	[0.0%]	[0.0%]	[0.0%]
inpatstays	Inpatient stays	[0.0%]	[0.0%]	[0.0%]
male	Gender: 1=male	[0.0%]	[0.0%]	[0.0%]
mammogram	Mammogram/breast X-ray in the last yr		[0.8%]	[12.9%]
maritals	Marital status	[0.0%]	[0.2%]	[1.8%]
nbrpeopl	Total number of people in household	[0.0%]	[0.2%]	
nbrrooms	# of rooms in home	[1.0%]	[0.6%]	
pap_smear	Pap smear in the last yr		[0.9%]	[13.5%]
pneushot	Pneumonia vaccination		[0.7%]	[13.1%]
povcat	Poverty Level Category	[8.3%]	[0.3%]	[80.8%]

Variable name	Label	NHANES	MCBS	MCBS Inst
prb_dres	Any difficulty dressing	[0.0%]	[0.7%]	[0.9%]
prb_eat	Any difficulty eating	[0.0%]	[0.7%]	[0.9%]
priv_insurr	Private Health Insurance	[0.0%]	[0.0%]	[0.0%]
priv_insurr	Private Health Insurance	[0.0%]	[0.0%]	[0.0%]
psa1yr	PSA test in last year		[2.6%]	[5.7%]
race	Race	[0.0%]	[0.3%]	[0.1%]
smokenow	Current smoker	[0.0%]	[0.4%]	[21.9%]
weightkg		[1.4%]	[0.7%]	[2.9%]

Table 1b. Prevalence(SE) [Rates of missingness] of SR variables available across the surveys(09)

Variable name	Label	NHANES	MCBS	MCBS Inst
cgarsr10	Cervical CA	0.60 (0.23) [0.1%]	0.46 (0.10) [0.0%]	
cgarsr11	Prostate CA	4.08 (0.61) [0.3%]	4.93 (0.30) [0.0%]	
cgarsr12	Hematological CA label	0.92 (0.37) [0.4%]		
cgarsr16	Diabetes	23.64 (1.39) [0.1%]	24.39 (0.60) [0.0%]	
cgarsr17	Undiagnosed Diabetes	2.42 (0.58) [0.1%]		
cgarsr18	Hyperlipidemia	51.45 (2.37) [2.9%]		
cgarsr19	Undiagnosed hyperlipidemia	5.35 (0.90) [2.9%]		
cgarsr37	Parkinson's Disease		1.59 (0.14) [0.0%]	15.28 (1.78) [19.4%]
cgarsr39	Paralysis		2.69 (0.20) [0.0%]	
cgarsr43	Cataract surgery		31.30 (0.67) [0.2%]	
cgarsr49	High blood pressure	63.58 (1.75) [0.1%]	69.07 (0.71) [0.1%]	
cgarsr50	Undiagnosed Hypertension	9.16 (1.35) [0.1%]		
cgarsr51	MI (heart attack)	8.57 (0.73) [0.4%]	13.58 (0.48) [0.0%]	
cgarsr52	Coronary Atherosclerosis and other heart disease		11.53 (0.45) [0.0%]	
cgarsr57	Stroke	8.17 (1.03) [0.3%]	11.24 (0.45) [0.0%]	
cgarsr6	Colon CA	2.71 (0.48) [0.4%]	2.49 (0.17) [0.0%]	
cgarsr67	COPD	9.85 (1.17) [0.3%]		
cgarsr68	Asthma	12.33 (1.46) [0.1%]		

Variable name	Label	NHANES	MCBS	MCBS Inst
cgarsr7	Lung CA	0.19 (0.09) [0.4%]	1.29 (0.14) [0.0%]	
cgarsr8	Skin cancer	12.27 (0.94) [0.4%]	22.20 (0.66) [0.0%]	
cgarsr82	Prostate Hyperplasia		9.48 (0.41) [0.8%]	
cgarsr88	Rheumatoid Arthritis		11.89 (0.48) [0.0%]	
cgarsr89	Arthritis: non-rheumatoid		57.78 (0.74) [0.0%]	
cgarsr9	Breast CA	5.68 (0.74) [0.1%]	5.09 (0.26) [0.0%]	
cgarsr91	Osteoporosis/brittle bones	15.79 (1.27) [0.6%]	22.98 (0.58) [0.0%]	
cgarsr96	Broken or fractured a hip	3.29 (0.57) [0.1%]	3.68 (0.18) [0.0%]	