

Supplementary Table 1: Operational coding of household water and sanitation exposures

Construct	NFHS variable	Classification	Codes / rule	Description
Drinking-water source	V113	Improved / no disadvantage	11, 12, 13, 14, 21, 31, 41, 51, 61, 62, 71, 92	Piped/public tap; tube well or borehole; protected well/spring; rainwater; tanker truck; cart with small tank; bottled water; community RO plant
Drinking-water source	V113	Unimproved / disadvantage	32, 42, 43, 96	Unprotected well; unprotected spring; surface water; other
Drinking-water source	V113	Unavailable	97	Treated as missing
Sanitation technology	V116	Improved technology	11, 12, 13, 15, 21, 22, 41	Flush to sewer, septic tank, or pit; flush destination unknown; VIP/biogas; pit with slab; twin-pit/composting
Sanitation technology	V116	Unimproved technology	14, 23, 31, 44, 96	Flush elsewhere; pit without slab/open pit; no facility/open defecation; dry toilet; other
Sanitation technology	V116	Unavailable	97	Treated as missing
Toilet sharing	V160	Not shared / shared / unavailable	0 / 1 / 7	Improved and shared = sanitation disadvantage; V160 not applicable for no facility/open defecation
Sanitation disadvantage	Derived	Disadvantaged	V116 unimproved OR V116 improved with V160=1	Study-specific indicator
Open defecation	Derived from V116	Yes	31	No facility/bush/field
Cumulative score	Derived	0, 1, or 2 disadvantages	Water disadvantage + sanitation disadvantage	Categorized as no, one, or two disadvantages

Note. Current JMP definitions classify packaged and delivered water as improved drinking-water sources, and classify shared improved sanitation facilities as a limited rather than basic sanitation service. [20] In this study, tanker-truck and cart-with-small-tank sources were therefore classified as improved water sources, while shared improved sanitation facilities were classified as sanitation disadvantage. V160 was structurally missing for respondents reporting no facility/open defecation; sharing status was used only when an improved sanitation technology was reported. RO = reverse osmosis; VIP = ventilated improved pit.

Supplementary Table 2: Direct comparison of respondents included in and excluded from the primary complete-case analysis

Characteristic	Included (n = 219,819), n (weighted %)	Excluded (n = 21,361), n (weighted %)	p-value
Age group			
15–19	112,701 (51.3%)	9,779 (44.6%)	<.001
20–24	107,118 (48.7%)	11,582 (55.4%)	
Residence			
Urban	48,990 (28.8%)	5,571 (34.8%)	<.001
Rural	170,829 (71.2%)	15,790 (65.2%)	
Household wealth			
Poorest	48,307 (20.1%)	4,647 (20.2%)	<.001
Poorer	52,549 (22.0%)	4,770 (19.9%)	
Middle	47,019 (21.2%)	4,256 (19.0%)	
Richer	40,879 (20.1%)	3,854 (19.2%)	
Richest	31,065 (16.5%)	3,834 (21.6%)	
Education			
No education	14,394 (6.4%)	1,616 (7.1%)	<.001
Primary	13,564 (6.1%)	1,518 (7.3%)	
Secondary	154,046 (68.7%)	14,086 (64.3%)	
Higher	37,815 (18.7%)	4,141 (21.3%)	
Marital status			
Never married	146,576 (64.7%)	11,648 (53.6%)	<.001
Married/in union	71,974 (34.8%)	9,583 (45.8%)	
Formerly married	1,269 (0.5%)	130 (0.6%)	
Number of living children			
0 children	170,647 (76.4%)	15,621 (72.4%)	<.001
1 child	30,445 (14.4%)	3,787 (18.1%)	
2+ children	18,727 (9.2%)	1,953 (9.5%)	
Pregnancy status			
Not pregnant	207,418 (94.2%)	19,593 (91.4%)	<.001
Currently pregnant	12,401 (5.8%)	1,768 (8.6%)	
Household size			
Small: <5	73,181 (33.7%)	5,583 (27.8%)	<.001
Medium: 5–7	108,785 (48.3%)	10,639 (48.7%)	
Large: 8+	37,853 (18.1%)	5,139 (23.5%)	
Media access			
No	44,826 (19.8%)	4,424 (19.7%)	.883
Yes	174,993 (80.2%)	16,937 (80.3%)	

Note. Counts are unweighted; percentages are survey-weighted column percentages. P-values were estimated using Rao-Scott survey-adjusted tests comparing respondents included in and excluded from the primary complete-case analysis. Inclusion required complete data for the primary outcome, both water and sanitation exposures, the BMI-derived indicator, model covariates, and survey-design variables. These comparisons characterize observed differences between groups but do not prove a missing-completely-at-random mechanism. BMI= body mass index.

Supplementary Table 3: Survey-weighted predictors of exclusion from the primary complete-case analysis

Characteristic	Adjusted OR (95% CI)	p-value
Age 20–24 years	1.04 (0.98, 1.10)	.178
Rural residence	0.72 (0.67, 0.76)	<.001
Wealth: Poorer	0.85 (0.81, 0.90)	<.001
Wealth: Middle	0.79 (0.74, 0.85)	<.001
Wealth: Richer	0.79 (0.72, 0.85)	<.001
Wealth: Richest	1.00 (0.91, 1.10)	.954
Education: Primary	1.17 (1.06, 1.28)	.001
Education: Secondary	1.04 (0.96, 1.13)	.324
Education: Higher	1.21 (1.09, 1.34)	<.001
Married or in union	2.04 (1.91, 2.17)	<.001
Formerly married	1.70 (1.33, 2.17)	<.001
One living child	0.80 (0.74, 0.85)	<.001
Two or more living children	0.60 (0.55, 0.65)	<.001
Currently pregnant	1.02 (0.95, 1.09)	.675
Household size: 5–7 members	1.28 (1.21, 1.35)	<.001
Household size: 8 or more members	1.56 (1.46, 1.66)	<.001
Media access	1.04 (0.99, 1.10)	.116

Note. Adjusted odds ratios represent the odds of exclusion from the primary complete-case analysis among all 241,180 age-eligible respondents. The model accounted for survey weights, clustering, and stratification and included state fixed effects. Reference categories were age 15–19 years, urban residence, poorest wealth quintile, no education, never married, no living children, not currently pregnant, household size below five, and no media access. Associations with observed characteristics provide evidence against missing completely at random but do not establish missing at random.