



Association of Household Water and Sanitation Disadvantage and Open Defecation with Moderate-to-Severe Anaemia Among Young Women in India: A Secondary Analysis of NFHS-5

Kinjal Nayak

Kent State University, Ohio, United States

Email: drkinjalnayak@gmail.com

Abstract

Introduction: Anaemia remains highly prevalent among young women in India, and water and sanitation conditions are biologically plausible contributors through faecal exposure and helminth-related blood loss. Among the studies identified, we found no NFHS-5 analysis of women aged 15–24 years that examined moderate-to-severe Anaemia using cumulative water/sanitation disadvantage score and a separate household open-defecation measure. This analysis was conducted to evaluate associations of cumulative household water/sanitation disadvantage and household open defecation with moderate-to-severe Anaemia. **Methods:** This cross-sectional analysis used NFHS-5 (2019–2021) data for women aged 15–24 years. The primary complete-case sample included 219,819 women, and multiple imputation included 232,082 women with observed water and sanitation exposures. The cumulative score combined drinking-water disadvantage with sanitation disadvantage, defined as unimproved sanitation technology or improved facility shared with one or more other households. Survey-weighted logistic regression adjusted for prespecified covariates and state fixed effects. **Results:** In the primary complete-case sample, the survey-weighted prevalence of moderate-to-severe Anaemia was 31.71%. The cumulative score was associated with the outcome overall ($p=0.002$). One disadvantage was associated with higher adjusted odds (AOR 1.06, 95%CI: 1.02, 1.09), whereas two disadvantages were not statistically significant (AOR 1.07, 95%CI: 0.98, 1.16). Sanitation disadvantage (AOR 1.06, 95%CI: 1.03, 1.10) and open defecation (AOR 1.08, 95%CI: 1.04, 1.12) were positively associated, while drinking-water disadvantage was not. Multiple-imputation findings were nearly identical, and no statistically significant interactions were observed, either before or after multiplicity correction. **Conclusion:** Sanitation disadvantage and open defecation were associated with modestly higher adjusted odds of moderate-to-severe Anaemia.

Key Words: Anemia, Defecation, Female, Sanitation, Water Supply

Introduction

Anaemia remains a public health concern among adolescent girls, young women, and women of reproductive age in India.^[1–3] NFHS-5 reported Anaemia among 57.0% of women aged 15 to 49 years, and Let et al.^[3] reported that Anaemia among women of reproductive age in India's Aspirational Districts increased from 58.7% in NFHS-4 to 61.1% in NFHS-5. Scott et al.^[2] reported Anaemia among 39.6% of Indian adolescent girls aged 10 to 19 years, with regional and state variation. Ibiroga et al.^[4] reported that Anaemia among Indian adolescents aged 15 to 19 years increased from 54.0% in NFHS-4 to approximately 59% in NFHS-5.

Anaemia is multifactorial and cannot be explained by iron deficiency alone.^[2,5] Studies of Indian adolescents and women have identified nutritional, genetic, inflammatory, socioeconomic, dietary, reproductive, and environmental contributors to Anaemia.^[2,5–7] Scott et al.^[2]

identified iron deficiency, hemoglobinopathies, vitamin A deficiency, and zinc deficiency as correlates of Anaemia among Indian adolescents. Gupta et al.^[5] found that Anaemia among rural adolescent girls in Haryana included iron deficiency, folate or vitamin B12 deficiency, mixed deficiencies, inflammation, and other causes. Kumar and Mohanty^[7] examined dietary diversity, hygiene behaviour, and socioeconomic factors in relation to adolescent undernutrition and Anaemia in India. Jana et al.^[6] reported that, in adjusted analyses, open defecation was associated with higher odds of Anaemia among women in West Bengal, while groundwater use was associated with higher odds of Anaemia in both West Bengal and Bangladesh; other associated factors differed between the two settings.

Women aged 15–24 years span age groups relevant to India's Anaemia-control strategy, including adolescents aged 10–19 years and women of reproductive age.^[8,9] The Anaemia Mukht Bharat framework described in the

2018 operational guidelines included adolescents and women of reproductive age among its target groups and incorporated interventions such as iron and folic acid supplementation, deworming, behaviour change communication, testing and treatment, fortified foods, and attention to selected non-nutritional causes of Anaemia.^[8] NFHS-5 is suitable for studying this age group because it collected health, nutrition, Anaemia, and reproductive information from women aged 15 to 49 years, while menstrual hygiene information was collected for women aged 15 to 24 years.^[1]

Water, sanitation, and hygiene conditions may contribute to Anaemia through faecal exposure, soil-transmitted helminth infection, hookworm-related intestinal blood loss, and impaired nutrient absorption.^[10-12] WHO states that soil-transmitted helminths are transmitted by eggs in human faeces that contaminate soil where sanitation is poor, and hookworms can cause chronic intestinal blood loss resulting in Anaemia, especially among adolescent girls and women of reproductive age.^[10] Studies have examined WASH-related exposures and Anaemia-related risk factors, including worm infestation, handwashing and footwear practices, open defecation, groundwater use, drinking-water and sanitation conditions, and community open defecation.^[6,12-14] Jana et al.^[6] reported that open defecation was associated with Anaemia among women in West Bengal, while groundwater use was associated with Anaemia among women in both West Bengal and Bangladesh. Chakrabarti et al.^[12] reported higher Anaemia and micronutrient-deficiency prevalence in higher community open-defecation settings among Indian children and adolescents, and another study by Chakrabarti et al.^[15] found that reduced village-level open defecation was associated with Anaemia reduction among pregnant Indian women.

Prior studies have examined Anaemia among adolescents, pregnant women, and women of reproductive age.^[2,6,7,12-15] Within this literature, WASH-related exposures and related factors have included sanitation conditions, open defecation, handwashing and footwear practices, worm infestation or deworming, groundwater and drinking-water source, toilet-facility characteristics, and community open defecation.^[6,7,12-15] For example, Kothari et al.^[11] analysed unimproved water and sanitation in relation to Anaemia across 47 nationally representative demographic and health surveys, Jana et al.^[6] examined women's Anaemia and open defecation in West Bengal and Bangladesh, and Chakrabarti et al.^[12] examined community open defecation, deworming, and Anaemia or micronutrient status among Indian children and adolescents. Recent nationally representative NFHS-5 research has examined drinking-water and

toilet-facility characteristics among a broader set of determinants of Anaemia in women aged 15–49 years.^[14] Closely related national studies have examined overlapping components of this question, Sunuwar et al.^[16] linked unimproved drinking water and sanitation with Anaemia among Indian women using NFHS-4; Sappani et al.^[17] examined moderate and severe Anaemia across NFHS-3 to NFHS-5; and Pedgaonker et al.^[18] evaluated household sanitation and community open defecation in relation to Anaemia using repeated NFHS data. These studies establish important associations involving Anaemia, sanitation, drinking water, and open defecation, but differ from the present study in population, Anaemia outcome, exposure definition, or level of analysis.

However, among the studies identified, we found no nationally representative NFHS-5 analysis focused on women aged 15 to 24 years that used moderate-to-severe Anaemia as the primary outcome while distinguishing a two-component household water/sanitation disadvantage score from household open defecation. To address this focused gap, this study used NFHS-5 data to examine the association between cumulative household water/sanitation disadvantage and moderate-to-severe Anaemia among young women aged 15 to 24 years in India. The analysis examined open defecation separately and assessed robustness through sensitivity analyses.

Methods

Study Design and Data Source:

This cross-sectional study used the India National Family Health Survey, fifth round (NFHS-5), conducted during 2019 to 2021.^[1] The analysis used the women's Individual Recode (IR) file, IAIR7EFL.sas7bdat, which contains each eligible woman's questionnaire data together with selected household variables. No separate Household Recode file was merged because the required household water, sanitation, socioeconomic, and survey-design variables were already present in the IR file. NFHS-5 used a multistage stratified sampling design, and the analysis accounted for the women's sampling weight (V005), primary sampling unit (V021), and sample stratum (V022), in accordance with DHS guidance.^[1,19] V022 and V023 contained identical values for all 241,180 age-eligible respondents; V022 was used as the stratification variable.

Study Population:

The study population included women aged 15 to 24 years who participated in NFHS-5. The initial age-restricted dataset included 241,180 women. A synchronized set of special codes for drinking water,

sanitation, and toilet sharing (V113=97, V116=97, and V160=7) occurred in 9,098 records and was treated as unavailable for this analysis rather than as an exposure category. This left 232,082 women with observed water and sanitation exposure data. Among these women, 2,640 were missing Anaemia information only, 485 were missing BMI only, and 9,138 were missing both Anaemia and BMI; the other model covariates and survey-design variables were complete. The primary analysis used complete cases with no missing outcome, exposures, BMI-derived indicator, covariates, and survey-design information and included 219,819 women. Multiple imputation was conducted as a sensitivity analysis among all 232,082 women with observed water and sanitation exposures. The unavailable water and sanitation exposures were not imputed.

Outcome Variables:

The primary outcome was moderate-to-severe Anaemia, derived from the NFHS Anaemia-level variable V457, which is based on haemoglobin measurement and NFHS adjustment procedures.^[1,19] V457 codes 1 (severe) and 2 (moderate) were coded as the outcome, while codes 3 (mild) and 4 (not anaemic) were coded as no moderate-to-severe Anaemia. Sensitivity outcomes were any Anaemia (V457 codes 1–3 versus 4) and severe Anaemia (V457 code 1 versus codes 2–4).^[1,19]

Exposure Variables:

The main cumulative exposure was a two-component household water/sanitation disadvantage score based on drinking-water source and sanitation conditions. Drinking-water sources were classified using a study-specific indicator informed by DHS and WHO/UNICEF Joint Monitoring Programme categories.^[19,20] Piped water, public taps, tube wells or boreholes, protected wells, protected springs, rainwater, tanker trucks, carts with small tanks, bottled water, and community reverse-osmosis plants were classified as improved. Unprotected wells, unprotected springs, surface water, and other sources were classified as unimproved.^[1,19,20]

Sanitation technology was classified as improved for flush or pour-flush facilities connected to a sewer, septic tank, or pit; flush facilities with an unknown destination; ventilated improved-pit or biogas latrines; pit latrines with a slab; and twin-pit or composting toilets. Flush facilities discharging elsewhere, pit latrines without a slab or open pits, no facility or open defecation, dry toilets, and other facilities were classified as unimproved. Sanitation disadvantage was defined as either an unimproved sanitation technology or an improved facility shared with one or more other households. Thus, shared improved facilities were retained as improved technology but classified as

disadvantaged because they represent a limited rather than basic sanitation service.^[20] Open defecation, identified from the no-facility/bush/field category, was evaluated separately.

The cumulative score summed drinking-water disadvantage and sanitation disadvantage and was categorized as no disadvantage, one disadvantage, or two disadvantages. These indicators reflect reported household source, facility technology, and sharing status and do not directly measure water quality, reliability, storage, toilet functionality or use, handwashing behaviour, or faecal exposure. Exact NFHS variable codes and classification rules are provided in [Supplementary Table 1](#).

Covariates:

Adjusted models included age group (15–19 or 20–24 years), urban or rural residence, household wealth quintile, educational level, marital status, number of living children, current pregnancy, household size, media access, BMI below 18.5 kg/m², and state fixed effects. These covariates were selected a priori based on their relevance to Anaemia, water and sanitation conditions, reproductive status, nutritional status, and socioeconomic patterning in NFHS-5 and prior literature.^[1,2,5-7,11,12] BMI was calculated as V445 divided by 100.^[19] Values corresponding to BMI below 10.00 or above 60.00 kg/m² were treated as missing. BMI below 18.5 kg/m² was used as an adult-threshold indicator; no adolescent-specific BMI-for-age classification was applied.

Statistical Analysis:

All analyses accounted for the NFHS complex survey design using the women's sampling weight (V005), primary sampling unit (V021), and sample stratum (V022). Sampling weights were divided by 1,000,000 before analysis, consistent with DHS guidance.^[19] Descriptive results are reported as unweighted counts with survey-weighted percentages. Differences by moderate-to-severe Anaemia status were evaluated using Rao-Scott survey-adjusted tests.^[21]

Survey-weighted logistic regression models used a quasibinomial family.^[21] The primary model estimated the association between the three-level cumulative water/sanitation disadvantage score and moderate-to-severe Anaemia, with no disadvantage as the reference group. A two-degree-of-freedom survey-adjusted Wald test evaluated the exposure overall.^[21] Secondary models examined drinking-water disadvantage, sanitation disadvantage, and open defecation separately. Because open defecation contributes to the sanitation-disadvantage construct, a model including both the cumulative score and open defecation was

treated as exploratory.

Sensitivity analyses used any Anaemia and severe Anaemia as alternative outcomes, restricted the sample to nonpregnant women, and restricted the sample to women with BMI of at least 18.5 kg/m². Included and excluded respondents were compared using unweighted counts and survey-weighted column percentages, with Rao-Scott survey-adjusted tests for differences in observed demographic and household characteristics. A survey-weighted logistic regression model of incomplete-case status then assessed the adjusted associations between observed characteristics and exclusion from the primary analysis. These analyses were used to characterize the missing-data pattern and were not interpreted as proving that data were missing completely at random. Multiple imputation was performed among 232,082 women with observed water and sanitation exposures. Twenty imputed datasets were generated over 10 iterations. Logistic regression was used to impute the binary Anaemia outcome, and predictive mean matching with five donors was used for BMI.^[22] The imputation models included the exposures, analysis covariates, state, and sampling weight; PSU and stratum identifiers were retained for survey analysis but were not entered as ordinary numeric predictors. Multiple imputation assumed that missing Anaemia and BMI values were missing at random conditional on the exposures, covariates, state, and sampling weight included in the imputation models; this assumption cannot be verified empirically.^[23] Survey-weighted models were fitted separately in each completed dataset and combined using Rubin's rules.^[23,24] Trace plots and the stability of outcome and BMI distributions across imputations were examined as diagnostics.

Exploratory interaction analyses tested effect modification by urban/rural residence and by a vulnerability group. The vulnerability score assigned one point each for being in the poorest or poorer wealth quintile, having no or primary education, BMI below 18.5 kg/m², having at least one living child, having no media access, and rural residence; scores of three or more defined higher vulnerability. Models involving vulnerability did not additionally adjust for the variables used to construct the score. Benjamini-Hochberg correction was applied across the four interaction tests.^[25] Adjusted generalized variance-inflation factors were examined for multicollinearity.^[26] An E-value was calculated for the open-defecation estimate using the nonrare-outcome conversion.^[27] Analyses were conducted in R version 4.6.1 using the survey, mice, mitools, car, and EValue packages.^[21,22,24,26,28]

Ethical Considerations:

The analysis used de-identified secondary NFHS-5 data

obtained from the DHS Program after authorization. No new data collection, participant contact, or identifiable information was involved. No separate institutional review board approval was obtained for this secondary analysis.

Results

Study population and analytic sample:

The age-restricted sample included 241,180 women aged 15–24 years. After excluding 9,098 women with unavailable water or sanitation exposure data, 232,082 women had observed exposure information. The primary complete-case sample included 219,819 women, representing 91.14% of the age-eligible sample; survey-weighted, 89.6% (95% CI: 89.3%, 89.8%) were included and 10.4% (95% CI: 10.2%, 10.7%) were excluded.

Included and excluded respondents differed across several observed characteristics, including age, residence, wealth, education, marital status, number of living children, pregnancy status, and household size, while media access was similar ([Supplementary Table 2](#)). In the adjusted missingness model, exclusion was more likely among women who were married or in union and those living in households with at least eight members, and less likely among rural residents ([Supplementary Table 3](#)). These patterns suggested that the complete-case sample differed from excluded respondents on observed characteristics; therefore, multiple imputation was conducted as a sensitivity analysis among all 232,082 women with observed water and sanitation exposures.

Descriptive characteristics and Anaemia prevalence:

Weighted characteristics of the complete-case sample are shown in Table 1. The survey-weighted prevalence of moderate-to-severe Anaemia was 31.71% (95% CI: 31.39%, 32.02%). Most women lived in rural areas, had secondary education, were never married, had no living children, were not currently pregnant, reported media access, and had BMI of at least 18.5 kg/m². Overall, most women had no water/sanitation disadvantage, while 20.0% reported open defecation.

Weighted distributions differed by Anaemia status for all characteristics shown in Table 1. Women with moderate-to-severe Anaemia had higher weighted proportions of water/sanitation disadvantage, open defecation, rural residence, lower wealth, lower educational attainment, having one or more living children, no media access, and BMI below 18.5 kg/m².

Adjusted associations between water, sanitation, and open-defecation indicators and moderate-to-severe Anaemia.

Table 1:
Unweighted counts and survey-weighted percentages of women aged 15–24 years by moderate-to-severe Anaemia status, NFHS-5 (N= 219,819)

Characteristic & Category	Overall (n = 219,819), n (weighted %)	No moderate-to-severe Anaemia (n = 150,093), n (weighted %)	Moderate-to-severe Anaemia (n = 69,726), n (weighted %)	p-value
Water/sanitation disadvantage				
No disadvantage	147,022 (67.1%)	102,387 (68.4%)	44,635 (64.3%)	<.001
One disadvantage	67,129 (31.1%)	44,084 (29.9%)	23,045 (33.7%)	
Two disadvantages	5,668 (1.8%)	3,622 (1.7%)	2,046 (2.1%)	
Open defecation				
No	178,874 (80.0%)	123,743 (81.0%)	55,131 (77.8%)	<.001
Yes	40,945 (20.0%)	26,350 (19.0%)	14,595 (22.2%)	
Age group				
15–19	112,701 (51.3%)	76,171 (50.6%)	36,530 (52.9%)	<.001
20–24	107,118 (48.7%)	73,922 (49.4%)	33,196 (47.1%)	
Residence				
Urban	48,990 (28.8%)	35,120 (29.9%)	13,870 (26.4%)	<.001
Rural	170,829 (71.2%)	114,973 (70.1%)	55,856 (73.6%)	
Household wealth				
Poorest	48,307 (20.1%)	31,061 (18.7%)	17,246 (23.2%)	<.001
Poorer	52,549 (22.0%)	35,583 (21.7%)	16,966 (22.8%)	
Middle	47,019 (21.2%)	32,202 (21.2%)	14,817 (21.2%)	
Richer	40,879 (20.1%)	28,563 (20.6%)	12,316 (19.0%)	
Richest	31,065 (16.5%)	22,684 (17.8%)	8,381 (13.7%)	
Education				
No education	14,394 (6.4%)	9,124 (6.0%)	5,270 (7.4%)	<.001
Primary	13,564 (6.1%)	8,727 (5.7%)	4,837 (6.9%)	
Secondary	154,046 (68.7%)	104,701 (68.2%)	49,345 (69.9%)	
Higher	37,815 (18.7%)	27,541 (20.1%)	10,274 (15.8%)	
Marital status				
Never married	146,576 (64.7%)	100,763 (65.3%)	45,813 (63.4%)	<.001
Married/in union	71,974 (34.8%)	48,469 (34.2%)	23,505 (36.1%)	
Formerly married	1,269 (0.5%)	861 (0.5%)	408 (0.5%)	
Number of living children				
0 children	170,647 (76.4%)	117,702 (77.3%)	52,945 (74.5%)	<.001
1 child	30,445 (14.4%)	20,351 (14.1%)	10,094 (15.2%)	
2+ children	18,727 (9.2%)	12,040 (8.6%)	6,687 (10.3%)	
Pregnancy status				
Not pregnant	207,418 (94.2%)	141,124 (93.9%)	66,294 (94.8%)	<.001
Currently pregnant	12,401 (5.8%)	8,969 (6.1%)	3,432 (5.2%)	
Household size				
Small: <5	73,181 (33.7%)	50,820 (34.1%)	22,361 (32.9%)	<.001
Medium: 5–7	108,785 (48.3%)	73,436 (47.9%)	35,349 (49.0%)	
Large: 8+	37,853 (18.1%)	25,837 (18.0%)	12,016 (18.1%)	
Media access				
No	44,826 (19.8%)	29,516 (19.0%)	15,310 (21.4%)	<.001
Yes	174,993 (80.2%)	120,577 (81.0%)	54,416 (78.6%)	
Underweight				
No	152,748 (68.1%)	106,591 (69.7%)	46,157 (64.6%)	<.001
Yes	67,071 (31.9%)	43,502 (30.3%)	23,569 (35.4%)	

Note. Counts are unweighted; percentages are survey-weighted column percentages. P-values were estimated using Rao-Scott survey-adjusted tests. BMI = body mass index; NFHS-5 = National Family Health Survey-fifth round.

Table 2:
Adjusted associations of household water and sanitation indicators with moderate-to-severe Anaemia

Model	Exposure	AOR (95% CI)	p-value	Overall p-value
Cumulative water/sanitation disadvantage	No disadvantage	1.00 (reference)	-	.002
	One disadvantage	1.06 (1.02, 1.09)	<.001	
	Two disadvantages	1.07 (0.98, 1.16)	.129	
Drinking-water model	Unimproved drinking-water source	1.00 (0.95, 1.06)	.903	
Sanitation model	Sanitation disadvantage	1.06 (1.03, 1.10)	<.001	
Open-defecation model	Open defecation	1.08 (1.04, 1.12)	<.001	

Note. All models used the primary complete-case sample (n=219,819), survey-weighted logistic regression with a quasibinomial family, and adjustment for age group, residence, wealth, education, marital status, number of living children, pregnancy, household size, media access, BMI below 18.5 kg/m², and state fixed effects. The overall p-value for the three-level cumulative score is from a two-degree-of-freedom survey-adjusted Wald test. AOR = adjusted odds ratio; CI = confidence interval

Table 3:
Sensitivity analyses of water/sanitation disadvantage, open defecation, and Anaemia outcomes

Analysis & Exposure	Sample, n	AOR (95% CI)	p-value	Overall score p-value
Any Anaemia				
One disadvantage	219,819	1.04 (1.01, 1.07)	.021	.022
Two disadvantages	219,819	1.085 (0.998, 1.180)	.057	
Open defecation	219,819	1.08 (1.04, 1.11)	<.001	
Severe Anaemia				
One disadvantage	219,819	1.19 (1.08, 1.30)	<.001	.001
Two disadvantages	219,819	1.05 (0.83, 1.34)	.668	
Open defecation	219,819	1.15 (1.04, 1.27)	.008	
Restricted to nonpregnant women				
One disadvantage	207,418	1.06 (1.02, 1.09)	.002	.005
Two disadvantages	207,418	1.08 (0.99, 1.17)	.089	
Open defecation	207,418	1.08 (1.04, 1.12)	<.001	
Restricted to women with BMI ≥18.5 kg/m ²				
One disadvantage	152,748	1.06 (1.02, 1.10)	.005	.012
Two disadvantages	152,748	1.09 (0.98, 1.20)	.101	
Open defecation	152,748	1.08 (1.03, 1.14)	.001	

Note. All models used the same covariate adjustment set as the primary model except when a restriction removed pregnancy or BMI from the analytic sample. The overall score p-value is from a two-degree-of-freedom survey-adjusted Wald test. The reference groups were no water/sanitation disadvantage and no open defecation. Three decimals are shown for the any-Anaemia estimate for two disadvantages because its lower confidence limit was close to 1.00. AOR = adjusted odds ratio; CI = confidence interval.

Table 4:
Complete-case retention and multiple-imputation sensitivity analysis

Component	Category or exposure	Sample, n	Estimate (95% CI)	p-value	FMI
Panel A. Complete-case retention					
Sample retention	Included in complete-case analysis	219,819	89.6% (89.3, 89.8)		
Sample retention	Excluded because of missing data	21,361	10.4% (10.2, 10.7)		
Panel B. Multiple-imputation sensitivity estimates					
Multiple imputation	One disadvantage	232,082	1.06 (1.02, 1.09)	<.001	4.0%
Multiple imputation	Two disadvantages	232,082	1.06 (0.98, 1.15)	.143	7.0%
Multiple imputation	Open defecation	232,082	1.08 (1.04, 1.12)	<.001	6.1%

Note. Panel A percentages and confidence intervals are survey weighted. Panel B estimates are adjusted odds ratios with 95% confidence intervals pooled across 20 imputed datasets using Rubin's rules. Water and sanitation exposures were observed and were not imputed. FMI = fraction of missing information.

Table 5:
Exploratory interaction analyses by residence and vulnerability group

Exposure	Potential modifier	Raw interaction p-value	BH-adjusted p-value
Water/sanitation disadvantage	Urban/rural residence	.940	.940
Open defecation	Urban/rural residence	.867	.940
Water/sanitation disadvantage	Vulnerability group	.847	.940
Open defecation	Vulnerability group	.113	.453

Note. All interaction models used the primary complete-case sample (n=219,819). Interaction terms were evaluated using survey-adjusted Wald tests. P-values were corrected across the four prespecified interaction tests using the Benjamini-Hochberg procedure. Higher vulnerability was defined as a score of at least three across household poverty, limited education, BMI below 18.5 kg/m², having one or more living children, no media access, and rural residence.

Adjusted associations are presented in Table 2. The cumulative water/sanitation disadvantage score was associated with moderate-to-severe Anaemia overall. Compared with no disadvantage, one disadvantage was associated with higher adjusted odds of moderate-to-severe Anaemia, whereas the estimate for two disadvantages was similar in magnitude but not statistically significant, providing no clear evidence of a monotonic dose-response pattern.

In separate models, drinking-water disadvantage was not associated with moderate-to-severe Anaemia. Sanitation disadvantage and open defecation were each associated with modestly higher adjusted odds. In an exploratory joint model including both the cumulative score and open defecation, the cumulative score was not significant overall, while open defecation remained modestly associated with the outcome. Adjusted generalized variance-inflation factors did not suggest problematic multicollinearity. The E-value for the open-defecation point estimate was modest.

Sensitivity, missing-data, and interaction analyses:

Sensitivity analyses were generally consistent with the primary findings (Table 3). One water/sanitation disadvantage and open defecation were associated with higher adjusted odds across alternative Anaemia outcomes and restricted-sample analyses, whereas estimates for two disadvantages were not statistically significant.

Multiple-imputation estimates were similar to the complete-case estimates (Table 4). Fractions of missing information were low. Trace plots showed overlapping chains without persistent drift, and imputed Anaemia and BMI distributions were stable across the 20 completed datasets.

There was no evidence that the associations differed by urban/rural residence or vulnerability group (Table 5). Raw interaction p-values ranged from .113 to .940, and all Benjamini-Hochberg-adjusted p-values exceeded .45.

Discussion

In this analysis of young women aged 15 to 24 years in India, the survey-weighted prevalence of moderate-to-severe Anaemia was 31.71%. This finding is consistent with prior evidence showing a high Anaemia burden among women and adolescents in India.^[1-4] NFHS-5 reported Anaemia among 57.0% of women aged 15 to 49 years; Scott et al.^[2] reported Anaemia among 39.6% of Indian adolescent girls aged 10 to 19 years; Ibirogba et al.^[4] reported that Anaemia among Indian adolescents aged 15 to 19 years increased from 54.0% in 2015 to 2016

to approximately 59% in 2019 to 2021; and Let et al.^[3] reported that Anaemia among women of reproductive age in India's Aspirational Districts increased from 58.7% in NFHS-4 to 61.1% in NFHS-5.^[1-4]

The cumulative household water/sanitation disadvantage score was associated with moderate-to-severe Anaemia overall after adjustment for demographic, socioeconomic, reproductive, household, nutritional, and state-level covariates (Table 2). One disadvantage was associated with approximately 6% higher adjusted odds, whereas the estimate for two disadvantages was similar in magnitude but imprecise and not statistically significant. This pattern does not provide clear evidence of a graded dose-response relationship. In separate models, sanitation disadvantage and open defecation were associated with modestly higher adjusted odds, while drinking-water disadvantage was not associated with the outcome.

The open defecation finding is compatible with biologically plausible pathways described in WASH, helminth, and Anaemia literature.^[10-12] WHO states that soil-transmitted helminths are transmitted through eggs in human faeces that contaminate soil where sanitation is poor, and that hookworms can cause chronic intestinal blood loss resulting in Anaemia, especially among adolescent girls and women of reproductive age.^[10] Prior India- and South Asia-specific studies have linked open defecation, groundwater use, worm infestation or deworming, and hygiene-related factors with Anaemia or related micronutrient outcomes.^[6,12,13,15] These findings support the relevance of sanitation-related pathways, although the present analysis did not directly measure helminth infection, faecal contamination, intestinal blood loss, inflammation, or micronutrient status.^[6,10,12,13,15]

The positive association observed for sanitation disadvantage is directionally consistent with Sunuwar et al.^[16], who reported higher adjusted odds of Anaemia among Indian women living in households with not-improved toilet facilities (aOR 1.14). That study also reported higher adjusted odds with not-improved drinking-water sources (aOR 1.12), whereas drinking-water disadvantage was not associated with moderate-to-severe Anaemia in the present study. Pedgaonker et al.^[18] found that improved non-shared household toilet facilities were associated with lower adjusted odds of co-occurring Anaemia among mother-child dyads in NFHS-4 and NFHS-5, while higher community-level open defecation was associated with higher adjusted odds across all three survey rounds. Differences in population, outcome definition, exposure measurement, and level of analysis limit direct comparison.

A recent nationally representative NFHS-5 analysis by Gnanasekaran et al.^[14] examined Anaemia among women

aged 15–49 years and included drinking-water source and toilet-facility characteristics among a broader set of sociodemographic, reproductive, behavioural, and environmental determinants. In that study, improved drinking-water sources were associated with slightly lower adjusted odds of Anaemia compared with unimproved sources (aOR= 0.97, 95% CI: 0.94, 0.99), and improved toilet facilities were associated with lower adjusted odds of Anaemia (aOR= 0.88, 95% CI: 0.87, 0.90). Their regression analyses modelled Anaemia as a binary anaemic-versus-not-anaemic outcome in the broader reproductive-age population. We did not identify separate analyses of household open-defecation status, shared improved sanitation as a sanitation-disadvantage construct, or a combined water/sanitation disadvantage count in that study. In contrast, the present study focused specifically on women aged 15–24 years and used moderate-to-severe Anaemia as the primary outcome. We observed no association for drinking-water disadvantage, whereas sanitation disadvantage and household open-defecation status were associated with modestly higher adjusted odds of moderate-to-severe Anaemia. Differences in age range, Anaemia outcome definition, WASH exposure construction, covariate specification, and analytic approach may contribute to the differing estimates across studies.^[14]

The modest and nonmonotonic pattern for the cumulative household water/sanitation disadvantage score may reflect differences between broad household source and facility indicators and more proximal exposure measures. Reported drinking-water source, sanitation technology, and facility sharing may not fully capture faecal exposure, pathogen burden, microbiological water contamination, water reliability, household water storage, sanitation use, toilet functionality, handwashing behaviour, or community-level sanitation conditions. Only 1.8% of the complete-case sample had two disadvantages, contributing to the wider confidence interval for that category. The null drinking-water estimate together with the positive sanitation and open-defecation estimates was more consistent with sanitation-related associations than with an association involving drinking-water source classification alone.

The modest size of water/sanitation-related associations is consistent with evidence that Anaemia is multifactorial.^[2,5,7,10] Prior studies have identified nutritional, genetic, inflammatory, socioeconomic, dietary, reproductive, water, sanitation, and hygiene-related factors associated with Anaemia or undernutrition among Indian adolescents, rural adolescent girls, and women in West Bengal and

Bangladesh.^[2,5-7] Thus, household water/sanitation disadvantage appears to be one component of a broader Anaemia risk profile rather than a dominant independent correlate in this analysis.^[2,5,7,10]

Descriptively, moderate-to-severe Anaemia was more common among young women living in rural areas, those in lower household wealth categories, those with no or primary education and a lower proportion with higher education, those with one or more living children, those with BMI below 18.5 kg/m², those with household water/sanitation disadvantage, and those living in households reporting open defecation (Table 1). However, the exploratory interaction analyses did not provide statistically significant evidence that the associations between household water/sanitation or open-defecation exposures and moderate-to-severe Anaemia differed by rural or urban residence or by the predefined vulnerability group after Benjamini-Hochberg correction (Table 5). These findings should be interpreted cautiously because the interaction analyses were exploratory and may not capture more granular differences in social, environmental, nutritional, or geographic vulnerability. Prior studies have shown that Anaemia risk is patterned by socioeconomic, nutritional, dietary, reproductive, and environmental factors, supporting the need for future analyses that examine more detailed vulnerability profiles.^[2,5-7]

These findings should be interpreted cautiously because the study was cross-sectional. The results support further investigation of open defecation as a marker of environmental and socioeconomic risk in Anaemia research among young women, particularly in analyses that also consider nutrition, deworming, household conditions, menstrual factors, infection, dietary intake, and broader environmental exposures. However, causal effects of open defecation or sanitation conditions on Anaemia cannot be inferred from these data.

Strengths and Limitations

This study contributes to existing WASH and Anaemia literature by distinguishing a cumulative household water/sanitation disadvantage score from open defecation in nationally representative NFHS-5 data for young women aged 15 to 24 years. Strengths include nationally representative data, biomarker-based haemoglobin measurement, focus on young women, adjustment for key covariates, and multiple-imputation sensitivity analysis.

Limitations include the cross-sectional design, residual confounding, primary reliance on complete cases, use of an adult BMI cutoff for respondents aged 15 to 17 years, and lack of adjustment for caste or social group. The

analysis incorporated whether an improved sanitation facility was shared, but it did not measure the number of sharing households, sanitation use or functionality, excreta management, water quality or reliability, handwashing, faecal exposure, helminth infection, inflammation, micronutrient status, menstrual factors, dietary intake, or local environmental contamination. In addition, only 1.8% of the complete-case sample had both water and sanitation disadvantages, resulting in limited precision for the two-disadvantage estimate and reducing our ability to assess a graded dose-response relationship. The household open-defecation indicator was derived from the reported no-facility/bush/field sanitation category and may therefore be subject to reporting or measurement bias. Any resulting exposure misclassification could affect the observed association, although the direction and magnitude of such bias cannot be determined from these data.

Conclusion

Moderate-to-severe Anaemia affected nearly one-third of young women aged 15 to 24 years in India. One household water or sanitation disadvantage, sanitation disadvantage, and open defecation were associated with modestly higher adjusted odds of moderate-to-severe Anaemia, whereas drinking-water disadvantage was not associated with the outcome. The estimate for two disadvantages was not statistically significant, and the findings did not demonstrate a clear dose-response relationship. Multiple-imputation results were consistent with the complete-case analysis. Because the study was cross-sectional and the E-value was modest, the associations should be interpreted cautiously.

Declaration

Funding: No funding was received for this research.

Conflict of Interest: No conflicts of interest are declared.

AI Tool Disclosure: OpenAI ChatGPT was used to assist with grammar and language.

References

1. International Institute for Population Sciences (IIPS) and ICF. National Family Health Survey (NFHS-5), 2019–21: India: Volume I. Mumbai: IIPS; 2021.
2. Scott S, Lahiri A, Sethi V, de Wagt A, Menon P, Yadav K, et al. Anaemia in Indians aged 10–19 years: prevalence, burden and associated factors at national and regional levels. *Matern Child Nutr*. 2022;18(4):e13391. doi:10.1111/mcn.13391
3. Let S, Tiwari S, Singh A, Chakrabarty M. Prevalence and determinants of anaemia among women of reproductive age in Aspirational Districts of India: an analysis of NFHS 4 and NFHS 5 data. *BMC Public Health*. 2024;24:437. doi:10.1186/s12889-024-17789-3
4. Ibirogba D, Menon VB, Olickal JJ, Thankappan KR. Trends in prevalence and predictors of Anaemia in adolescents between the ages of 15 and 19 years in India and its states: evidence from the National Family Health Survey 2015–16 and 2019–21. *Cureus*. 2024;16(10):e70733. doi:10.7759/cureus.70733
5. Gupta A, Sachdev HS, Kapil U, Prakash S, Pandey RM, Sati HC, et al. Characterisation of anaemia amongst school going adolescent girls in rural Haryana, India. *Public Health Nutr*. 2022;25(12):3499–3508. doi:10.1017/S1368980022000210
6. Jana A, Chattopadhyay A, Saha UR. Identifying risk factors in explaining women's anaemia in limited resource areas: evidence from West Bengal of India and Bangladesh. *BMC Public Health*. 2022;22:1433. doi:10.1186/s12889-022-13806-5
7. Kumar M, Mohanty PC. Undernutrition and anaemia among Indian adolescents: role of dietary diversity and hygiene practices. *J Nutr Sci*. 2023;12:e33. doi:10.1017/jns.2023.19
8. National Health Mission, Ministry of Health and Family Welfare, Government of India. Anaemia Mukh Bharat: Operational Guidelines for Programme Managers. New Delhi: Ministry of Health and Family Welfare, Government of India; 2018. <https://www.nhm.gov.in/images/pdf/Nutrition/AMB-guidelines/Anemia-Mukh-Bharat-Operational-Guidelines-FINAL.pdf>
9. Rizvi ZA, Singh J, Mishra P, Kumar A, Saini A, Patel N, et al. Anaemia Mukh Bharat Index: methodology and state rankings of iron and folic acid supplementation coverage in India, 2018–2019 to 2022–2023. *Glob Health Sci Pract*. 2025;13(2):e2400077. doi:10.9745/GHSP-D-24-00077
10. World Health Organization. Soil-transmitted helminth infections. World Health Organization; 2023. Accessed May 30, 2026. <https://www.who.int/news-room/fact-sheets/detail/soil-transmitted-helminth-infections>
11. Kothari MT, Coile A, Huestis A, Pullum T, Garrett D, Engmann C. Exploring associations between water, sanitation, and Anaemia through 47 nationally representative demographic and health surveys. *Ann N Y Acad Sci*. 2019;1450(1):249–267. doi:10.1111/nyas.14109
12. Chakrabarti S, Ajampur SSR, Waddington HS, Kishore A, Nguyen PH, Scott S. Deworming and micronutrient status by community open defecation prevalence: an observational study using nationally representative data from India, 2016–2018. *PLoS Med*. 2024;21(5):e1004402. doi:10.1371/journal.pmed.1004402

13. Siva PM, Sobha A, Manjula VD. Prevalence of anaemia and its associated risk factors among adolescent girls of central Kerala. *J Clin Diagn Res.* 2016;10(11):LC19-LC23. doi:10.7860/JCDDR/2016/20939.8938
14. Gnanasekaran S, Gupta MK, Jayaraj V, Mohan G, Rajendran V. Distribution and determinants of Anaemia among reproductive age group women in India: an insight from the fifth round of the National Family Health Survey (2019–21). *Indian J Community Med.* 2026;51(2):305–313. doi:10.4103/ijcm.ijcm_429_24.
15. Chakrabarti S, George N, Majumder M, Raykar N, Scott S. Identifying sociodemographic, programmatic and dietary drivers of anaemia reduction in pregnant Indian women over 10 years. *Public Health Nutr.* 2018;21(13):2424–2433. doi:10.1017/S1368980018000903
16. Sunuwar DR, Singh DR, Chaudhary NK, Pradhan PMS, Rai P, Tiwari K. Prevalence and factors associated with Anaemia among women of reproductive age in seven South and Southeast Asian countries: evidence from nationally representative surveys. *PLoS One.* 2020;15(8):e0236449. doi:10.1371/journal.pone.0236449
17. Sappani M, Mani T, Asirvatham ES, Joy M, Babu M, Jeyaseelan L. Trends in prevalence and determinants of severe and moderate anaemia among women of reproductive age during the last 15 years in India. *PLoS One.* 2023;18(6):e0286464. doi:10.1371/journal.pone.0286464
18. Pedgaonker S, Meher T, Gupta M, Chakrabarti S, Nguyen PH, Singh SK, et al. Anaemia among mother–child dyads in India: trends, drivers, and future projections. *Matern Child Nutr.* 2025:e70106. doi:10.1111/mcn.70106
19. Croft TN, Allen CK, Zachary BW, et al. Guide to DHS Statistics. Rockville, MD: ICF; 2023. https://dhsprogram.com/pubs/pdf/DHSG1/Guide_to_DHS_Statistics_DHS-8.pdf
20. World Health Organization, UNICEF. Progress on Household Drinking Water, Sanitation and Hygiene 2000–2020: Five Years Into the SDGs. Geneva: World Health Organization and United Nations Children's Fund; 2021.
21. Lumley T. Analysis of complex survey samples. *J Stat Softw.* 2004;9(8):1–19. doi:10.18637/jss.v009.i08
22. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw.* 2011;45(3):1–67. doi:10.18637/jss.v045.i03
23. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ.* 2009;338:b2393. doi:10.1136/bmj.b2393
24. Lumley T. mitools: Tools for Multiple Imputation of Missing Data. R package version 2.4. 2019. doi:10.32614/CRAN.package.mitools
25. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol.* 1995;57(1):289–300. doi:10.1111/j.2517-6161.1995.tb02031.x
26. Fox J, Weisberg S. *An R Companion to Applied Regression.* 3rd ed. Thousand Oaks, CA: Sage; 2019.
27. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. *Ann Intern Med.* 2017;167(4):268–274. doi:10.7326/M16-2607.
28. R Core Team. R: A Language and Environment for Statistical Computing. Version 4.6.1 [computer software]. Vienna, Austria: R Foundation for Statistical Computing; 2026. Accessed July 13, 2026. <https://www.R-project.org/>