Determinants of Anaemia in Pregnancy in India: A Comparative Weighted Logistic Regression Analysis Using NFHS-4 and NFHS-5

Sanjay Karande[1,2], R. L. Shinde[2]
[1,2]Department of Statistics, Sathaye College (Autonomous), Mumbai-400057, India.
[2]Department of Actuarial Science, KBC-North Maharashtra University, Jalgaon-42500, India.

ABSTRACT

Background: In India, the issue of anaemia during pregnancy, defined by low haemoglobin levels, continues as a major public health concern. Recent national surveys indicate a rising anaemia rate among pregnant women, regardless of large-scale nutritional programmes. This study aims to investigate the factors influencing anaemia by comparing data from NFHS-4 (2015–16) and NFHS-5 (2019–21) using a comprehensive analytical method that considers socioeconomic, demographic, nutritional, and environmental factors.

Methodology: Data were obtained from samples that represent the national population of women who are currently pregnant, as recorded in NFHS-4 and NFHS-5. Weighted logistic regression models, adjusted for the complex survey design, were used to estimate adjusted odds ratios (AORs) and 95% confidence intervals. The variables considered were maternal demographics, socioeconomic status, BMI, ANC usage, WASH conditions, and media exposure. The model's performance was checked using AIC, McFadden's pseudo-R^2, and ROC curves. Results were considered significant if p was less than 0.05.

Findings: More than half of the people had anaemia in both survey rounds. The characteristics of the population remained largely unchanged. BMI was the most reliable factor in preventing anaemia. Education helped prevent anaemia in NFHS-4 but was less effective in NFHS-5. Wealth became a bit more protective in the later survey. Differences related to caste were more noticeable in NFHS-5.ANC utilisation, WASH conditions, media exposure, and birth interval contributed slightly in adjusted models. Overall model discrimination remained moderate among rounds.

Conclusions: Maternal nutritional status continues to play a crucial role in anaemia risk, while socioeconomic and social-group factors are evolving. To effectively reduce anaemia during pregnancy, it is crucial to enhance nutrition-focused interventions and address structural inequities.

Keywords: Anaemia, Pregnancy, NFHS, Weighted Logistic Regression model.

1. INTRODUCTION

In India, anaemia during pregnancy is a significant health concern, impacting numerous women and leading to serious health complications for both mothers and their infants. According to the World Health Organisation, anaemia in pregnant women is defined by haemoglobin levels falling below 11 g/dL, which impairs the blood's ability to transport sufficient oxygen, thereby increasing vulnerability to health issues (WHO, 2011). Despite the introduction of national health initiatives and improvements in healthcare, India continues to have one of the highest rates of anaemia among pregnant women globally (Kalaivani & Ramachandran, 2018). This issue highlights a profound social and nutritional disparity affecting maternal health (IIPS & ICF, 2021).

Anaemia during pregnancy can lead to severe complications, including premature births, low birth weights, stillbirths, maternal infections, and fatigue that hampers daily activities (Kassebaum et al., 2014). The primary cause is often iron deficiency, but insufficient folate, vitamin B12, and vitamin A also hinder haemoglobin production (Molloy et al., 2008). Infections such as intestinal worms exacerbate the condition, particularly in rural and tribal regions with inadequate sanitation and healthcare (Sehgal et al., 2010).

The risk of anaemia in pregnant women in India is influenced by social and economic factors. Women with lower educational levels, reduced family income, and limited autonomy in health-related decisions are more susceptible to anaemia (Suryanarayana et al., 2016). Communities that

are marginalised, such as Scheduled Castes and Scheduled Tribes, face increased risks due to having fewer resources, insufficient dietary variety, and limited access to nutrition and prenatal care services (Debnath et al., 2021).In northeastern and tribal areas, distinctive dietary practices, high infection rates, and poor utilisation of antenatal services further elevate anaemia rates (Mog et al., 2023).

Diet and nutrition are important in preventing anaemia. Many pregnant women do not eat enough of iron-rich foods like pulses, fish, and leafy greens, which helps reduce the risk (Let et al., 2024). Low body weight, not eating enough calories, and long-term undernutrition link poverty to anaemia, causing poor health across generations (Bharati et al., 2008). Factors like access to health information, understanding supplements, and regular antenatal visits affect whether women take recommended iron and folic acid (Chakrabarti et al., 2018).

Environmental conditions also matter. Poor sanitation, dirty drinking water, and bad hygiene can lead to infections. These infections make it hard for the body to absorb nutrients and can increase the risk of anaemia, especially in rural and peri-urban areas (Patel et al., 2019). These environmental challenges combine with social and nutritional disadvantages, creating a complex set of risks for pregnant women.

Government programs such as the National Nutritional Anaemia Prophylaxis Programme, National Iron Plus Initiative, and Anaemia Mukt Bharat have worked to address anaemia through supplements, deworming, dietary advice, and behavior change efforts (Kapil et al., 2019). Recent data from the National Family Health Survey (NFHS-5) shows that anaemia rates have gone up compared to earlier surveys. This suggests there are gaps in program reach and changes in underlying factors (IIPS & ICF, 2021). Studies show that changes in behaviour and social and economic factors, not just population changes, explain much of this increase (Let et al., 2024).

Given this situation, it is important to study the factors affecting anaemia in pregnancy using data from NFHS-4 (2015–16) and NFHS-5 (2019–21). Weighted logistic regression is used to accurately compare survey rounds, considering the complex sampling methods. This study looks for both ongoing and new risk factors. It aims to offer updated evidence to help create better-targeted programs and enhance maternal nutrition programs in India.

2. METHODOLOGY

2.1 Study Design: This study used a cross-sectional comparative design based on data from two rounds of the National Family Health Survey (NFHS)—NFHS-4 (2015–16) and NFHS-5 (2019–21). The Ministry of Health and Family Welfare conducted both surveys, with the International Institute for Population Sciences (IIPS) serving as the central agency. They utilised the standard methods of the Demographic and Health Survey (DHS) to ensure the data accurately reflects the entire nation (IIPS & ICF, 2017; IIPS & ICF, 2021).

2.2 Data Source and Study Population: The analysis used the women's individual datasets from NFHS, which include detailed information on fertility, maternal health, nutrition, household environment, and haemoglobin levels. The study focused only on women who were pregnant at the time of the survey, identified through pregnancy-related questions. Once these women were chosen, distinct nationally representative samples were formed for NFHS-4 and NFHS-5. To ensure the integrity of the data, women with missing, invalid, or biologically improbable haemoglobin levels were excluded, in line with methods used in previous anaemia studies based on NFHS (Debnath et al., 2021; Let et al., 2024).

2.3 Outcome Variable: Anaemia status was the main outcome, measured using haemoglobin levels taken with a HemoCue analyser adjusted for altitude and smoking. According to WHO and NFHS guidelines, pregnant women with haemoglobin below 11.0 g/dL were classified as anaemic (including mild, moderate, and severe categories), while those with levels 11.0 g/dL or above were classified as not anaemic (WHO, 2011; IIPS & ICF, 2021).

2.4 Explanatory Variables: A set of demographics, socioeconomic, nutritional, environmental, and maternal health variables was chosen based on previous research and their availability in both NFHS rounds. To allow valid comparison, all variables were harmonised between NFHS-4 and

NFHS-5. Key factors considered were the mother's age category, level of education, wealth status, whether they lived in urban or rural areas, caste or tribal affiliation, maternal BMI, the number of antenatal care (ANC) appointments attended, the count of children under five years old, and two composite indices:

- WASH index, combining water source, sanitation, flooring quality, and hygiene facilities
- Media exposure index, combining frequency of reading newspapers, watching television, and listening to the radio

These combined measures help avoid repeating similar information and cover a wider range of environmental and informational factors, using standard public health modelling methods (Chakrabarti et al., 2018; Mog et al., 2023).

2.5 Survey Weights and Complex Design Adjustment: Given that the NFHS employs a stratified, two-stage cluster sampling design, all analyses incorporated sample weights and accounted for the survey's clustering and stratification. The application of weights and design correction ensures that the results are nationally representative and account for variations in selection probability and non-response (IIPS & ICF, 2021).

2.6 Statistical Analysis: Separate weighted logistic regression analyses were conducted for NFHS-4 and NFHS-5 to determine the factors linked to anaemia during pregnancy. These models calculated adjusted odds ratios (AORs) along with 95% confidence intervals for each variable. To address the issue of overdispersion often found in extensive survey data, a quasibinomial model was employed (Chakrabarti et al., 2018).

To diagnose the model, variance inflation factors were used to check for multicollinearity, the Akaike Information Criterion (AIC) and McFadden's pseudo R^2 were employed to assess model fit, and receiver operating characteristic (ROC) curves, along with the area under the curve (AUC), were utilised to evaluate discrimination. Ultimately, AORs from both survey rounds were compared to investigate changes in the determinants of anaemia over time.

All analyses were performed in R using the survey package to account for the complex survey design.

3. RESULTS

3.1 Anaemia prevalence and sample characteristics: The final weighted samples included 1,859 pregnant women from NFHS-4 and 3,536 from NFHS-5 (Table 1). In both survey rounds, over half of the pregnant women were anaemic, reflecting a persistently high burden. Demographic and socioeconomic profiles, such as age distribution, education levels, wealth quintiles, and rural–urban residence, were broadly similar across the two rounds, allowing for meaningful comparison of anaemia determinants.

3.2 Model performance: The evaluation metrics in Table 2 show that the NFHS-4 model fits better overall. It has a McFadden's pseudo-R^2 of 0.5415, while NFHS-5 has 0.0274. NFHS-4 also has lower AIC and BIC values, which means it explains the data better. The ROC curves in Figures 1 and 2 show that both models have modest discrimination ability, with AUC values of 0.596 for NFHS-4 and 0.584 for NFHS-5. Table 1 shows that Sensitivity and specificity are slightly higher in NFHS-4. According to the multicollinearity diagnostics presented in Table 4, the variance inflation factors for all predictors are low, suggesting that the model's estimates remain consistent.

3.3 Determinants of anaemia :

Maternal age: Table 3 shows that older pregnant women had a slightly lower chance of having anaemia in both rounds (NFHS-4 OR = 0.97; NFHS-5 OR = 0.98), but this difference was not significant.

Education: Education was a significant protective factor in NFHS-4 (OR = 0.71), but this effect weakened and became non-significant in NFHS-5 (OR = 0.91). This change suggests that the benefits of education on maternal nutrition or health behaviours may have diminished over time (Table 3; Figures 3 and 4).

Wealth index: Wealth showed a positive but non-significant association with anaemia in NFHS-4 (OR = 1.23), which reversed to a mildly protective, though non-significant, association in NFHS-5

(OR = 0.93). This shift may reflect evolving socioeconomic influences on diet quality and supplement access (Table 3).

Caste/tribe group: In NFHS-4, belonging to a specific caste or tribe slightly decreased the risk (OR = 0.94). However, in NFHS-5, it slightly increased the risk (OR = 1.14). This indicates that social disadvantage may now hold greater significance (Table 3).

Place of residence: In NFHS-4, living in urban areas seemed to offer a small, non-significant protective effect (OR = 0.73). But in NFHS-5, it appeared to slightly increase risk, though not significantly (OR = 1.11). This change might be due to shifts in urban eating habits and ways of living (Table 3).

Regional variation: Region or state showed minimal effects in both rounds, with odds ratios near unity, implying little geographic variation after adjusting for other factors (Table 3).

Body Mass Index (BMI) : BMI was the most consistent and significant protective factor in both surveys (NFHS-4 OR = 0.96; NFHS-5 OR = 0.96), indicating that higher maternal BMI substantially lowers anaemia risk during pregnancy (Figures 3 and 4).

Maternal healthcare utilisation: The number of antenatal care visits and tetanus injections had odds ratios close to 1.00,&were not significantly associated with anaemia in either survey round (Table 3).

Reproductive history: Parity was significantly associated with increased anaemia risk in NFHS-4 (OR = 1.34), but this association weakened and became non-significant in NFHS-5 (OR = 1.05), possibly reflecting improvements in nutritional support or birth spacing (Table 3).

Birth interval: Birth interval showed no meaningful association with anaemia in either survey round (Table3).

WASH and media exposure indices: The combined scores for water, sanitation, hygiene (WASH), and media exposure had small and not significant links to anaemia in both rounds. This means they did not have much direct effect when other factors were considered (Table 3).

3.4 Comparative trends across NFHS-4 and NFHS-5 : Looking at the adjusted odds ratios from the two survey rounds (Figures 3–5), we see some important changes over time. Education's protective effect got weaker. Wealth now has a slight protective effect in NFHS-5. Social differences related to caste or tribe became more noticeable. BMI stayed the strongest and most stable protective factor. The link between having more children and anaemia got weaker, which might mean better maternal health practices.

These patterns show that while good nutrition is still important, the role of social and economic factors in anaemia risk is changing. This means public health plans need to adapt.

4. DISCUSSION

This comparative analysis of NFHS-4 and NFHS-5 highlights both persistent and shifting determinants of anaemia in pregnancy in India. The most consistent finding across rounds was the strong protective association of maternal BMI, reaffirming the central role of women’s nutritional status in shaping anaemia risk. Previous studies have similarly shown that chronic energy deficiency and inadequate micronutrient intake are major contributors to maternal anaemia in India (Chakrabarti et al.; Mog et al.). The stability of BMI’s effect across time underlines the need for sustained nutritional interventions before and during pregnancy.

Socioeconomic factors exhibited more variation. Education, which significantly reduced anaemia odds in NFHS-4, showed a weakened effect in NFHS-5. This suggests that while education traditionally enhances awareness of dietary practices and supplementation, broader constraints—such as dietary affordability, limited diversity, or inconsistent counselling—may hinder its effectiveness (Let et al.; Debnath et al.). Likewise, the wealth index shifted from a neutral association in NFHS-4 to a mildly protective pattern in NFHS-5, aligning with evidence that improved economic status increasingly influences diet quality and supplement adherence (Mog et al.).

The influence of caste/tribe became more pronounced in NFHS-5, reflecting persistent social inequalities that affect food access, health service utilisation, and overall nutritional status (Debnath

et al.). Meanwhile, indicators such as ANC visits, tetanus injections, WASH conditions, media exposure, and birth spacing showed minimal independent associations, consistent with findings that behavioural and environmental factors often exert indirect rather than direct effects on anaemia (Chakrabarti et al.).

Overall, findings emphasise that while nutritional determinants remain central, socioeconomic and social-group inequalities are evolving, underscoring the need for integrated, equity-focused maternal nutrition strategies.

5. CONCLUSION

This study looks at what causes anaemia in pregnant women in India. It compares two large national datasets, NFHS-4 and NFHS-5. The study uses harmonised variables and weighted logistic regression to show how nutrition, socioeconomic status, and demographics affect the risk of maternal anaemia over time. This approach is good at showing changes in how education, wealth, and social groups affect things. It also shows that nutritional measures like BMI stay the same. This comparison over time shows that anaemia during pregnancy is linked to both biological factors and changing social conditions.

Anaemia is not caused by just one thing. It is linked to reflect the intersection of nutritional adequacy, economic capability, social position, and broader household environments.

The findings highlight that reducing anaemia should not be seen as a separate medical goal. It should be part of a larger plan for improving mothers' health and nutrition. Improving dietary diversity, strengthening pre-pregnancy nutrition, and ensuring continuity of iron and folic acid intake are essential components. Equally important is addressing the uneven benefits of education and economic improvements, which appear to translate differently across contexts and over time. This study compares NFHS-4 and NFHS-5 to highlight the need for policies that adapt to changing social and economic trends while still focusing on important nutritional support. Continuous monitoring and the implementation of targeted, equity-focused programs are crucial for expeditiously advancing the reduction of anaemia among pregnant women in India.

Table 1. Model performance: confusion matrix

NFHS-4				NFHS-5			
Actual				Actual			
Predicted	0	1	Total	Predicted	0	1	Total
0	117	100	217	0	548	417	965
1	655	987	1,642	1	1,110	1,461	2,571
Total	772	1,087	1,859	Total	1,658	1,878	3,536

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Table 2. Model performance statistics

	NFHS-4	**NFHS-5**
Metric	**Value**	**Value**
Pseudo-R^2 (McFadden)	0.5415	0.0274
AIC	1185.00	4782.36
BIC	1251.72	4868.75
AUC (ROC)	0.5960	0.584
95% CI for AUC	0.5581 – 0.6338	0.5649 – 0.6025

Table 3. Adjusted odds ratios for anaemia among pregnant women

	NFHS-4		NFHS-5

Predictor	OR	95% CI Lower	95% CI Upper		OR	95% CI Lower	95% CI Upper
Age (years)	0.97	0.91	1.02		0.98	0.95	1.01
Education level	0.71	0.57	0.88		0.91	0.82	1.02
Wealth index	1.23	0.99	1.53		0.93	0.85	1.02
Caste/Tribe group	0.94	0.73	1.21		1.14	0.99	1.31
Residence (urban/rural)	0.73	0.45	1.20		1.11	0.86	1.45
Region/State	0.99	0.94	1.04		1.01	1.00	1.02
BMI (kg/m²)	0.96	0.95	0.98		0.96	0.96	0.97
ANC visits	1.01	0.99	1.03		0.99	0.99	1.00
Tetanus injections	1.00	1.00	1.00		1.00	1.00	1.00
Children ever born	1.34	1.06	1.69		1.05	0.94	1.16
Birth interval (months)	1.01	0.99	1.02		1.00	1.00	1.01
Media exposure index	0.96	0.84	1.09		1.03	0.95	1.12
WASH index	1.00	1.00	1.01		1.00	1.00	1.00

Table 4. Multicollinearity diagnostics (VIF)

	NFHS-4		NFHS-5
Variable	**VIF**		**VIF**
Age (years)	1.46		1.50
Education level	1.29		1.26
Wealth index	2.06		1.60
Caste/Tribe group	1.12		1.07
Residence	1.22		1.21
Region/State	1.18		1.06
BMI	1.04		1.02
ANC visits	1.04		1.06
Tetanus injections	1.03		1.07
Children ever born	1.43		1.46
Birth interval	1.13		1.17
Media exposure index	2.78		2.95
WASH index	2.83		2.91

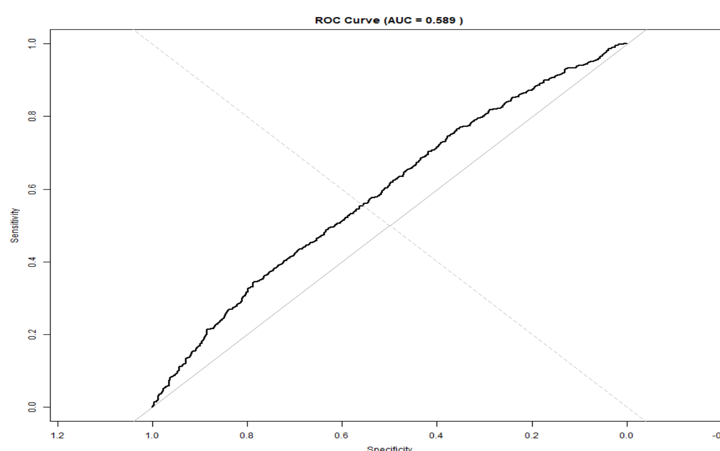

Figure 1. ROC curve for weighted logistic regression model predicting anaemia among pregnant women, NFHS-4 (AUC = 0.596).

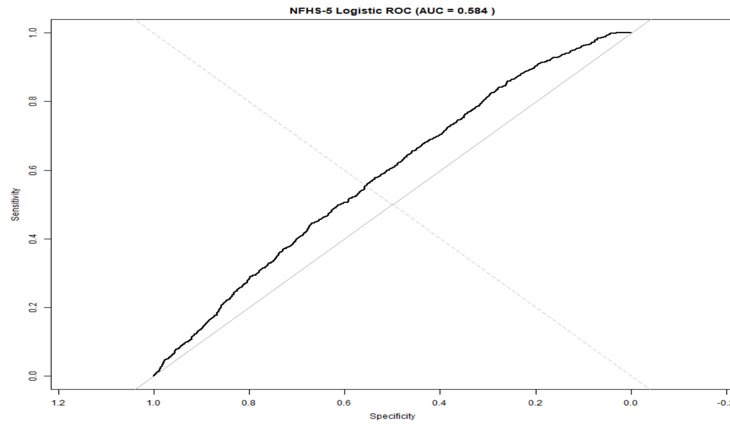

Figure 2. ROC curve for weighted logistic regression model predicting anaemia among pregnant women, NFHS-5 (AUC = 0.584).

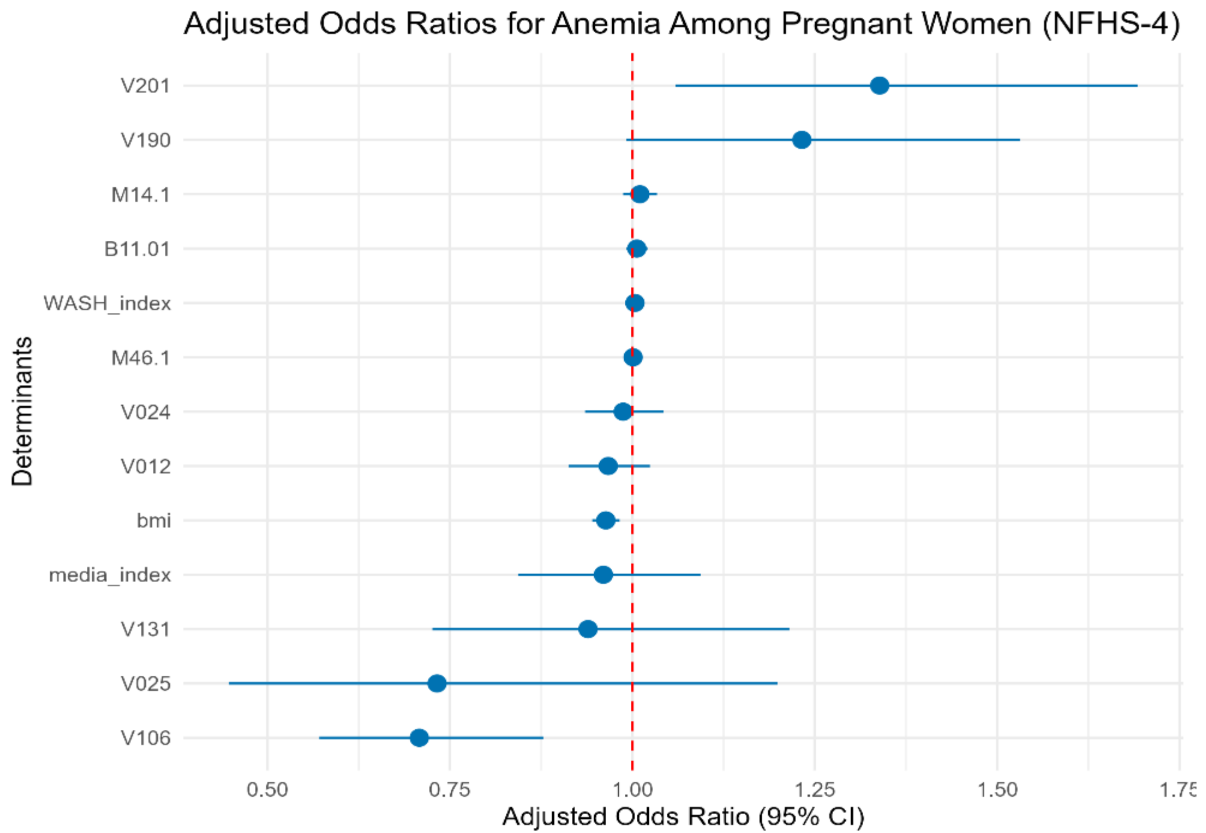

Figure 3. Adjusted odds ratios (95% CI) for determinants of anaemia among pregnant women, NFHS-4.

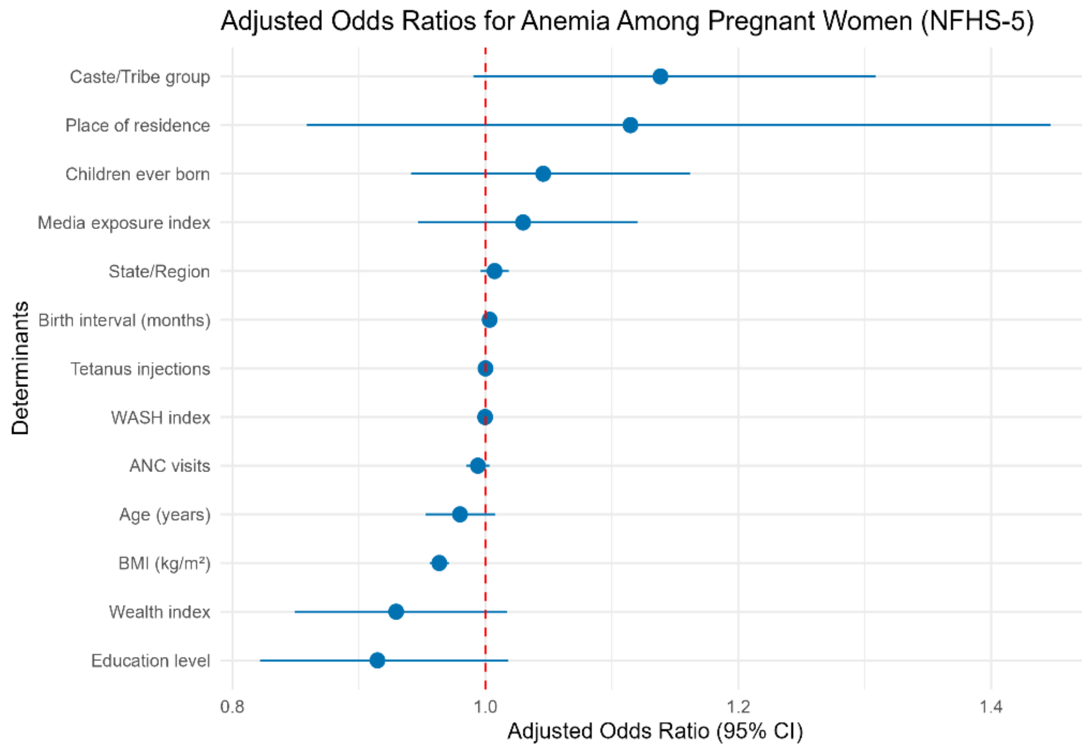

Figure 4. Adjusted odds ratios (95% CI) for determinants of anemia among pregnant women, NFHS-5.

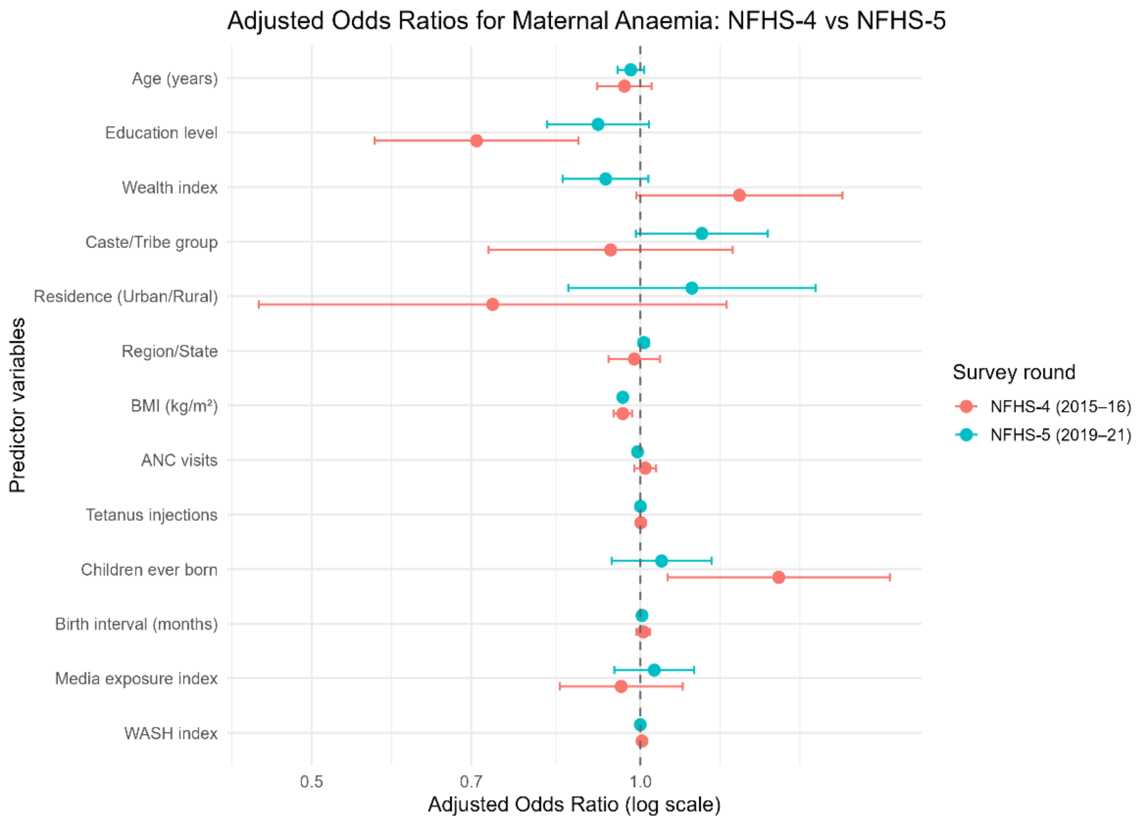

Figure 5. Comparison of adjusted odds ratios for maternal anaemia between NFHS-4 and NFHS-5 (pooled forest plot).

ACKNOWLEDGEMENT

The author expresses gratitude to KBC North Maharashtra University for the Registration for the PhD Degree, my Guide and Family members.

Declaration of competing interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
Data Availability Statement: The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

REFERENCES

1. Bharati P, Som S, Chakrabarty S, Bharati S, Pal M. Prevalence of anaemia and its determinants among nonpregnant and pregnant women in India. *Asia Pac J Public Health.* 2008;20(4):347–359. https://doi.org/10.1177/1010539508322762.
2. Chakrabarti S, George N, Majumder M, Raykar N, Scott S. Drivers of anaemia reduction in pregnant Indian women. *Public Health Nutr.* 2018;21(5):893–904.
3. Debnath A, Debbarma A, Debbarma SK, Bhattacharjya H. Factors associated with anaemia in antenatal clinic attendees in Northeast India. *J Family Med Prim Care.* 2021;10(3):1106–1111.
4. IIPS & ICF. *National Family Health Survey (NFHS-4), India Report.* Mumbai: International Institute for Population Sciences; 2017.
5. IIPS & ICF. *National Family Health Survey (NFHS-5), India Report.* Mumbai: International Institute for Population Sciences; 2021.
6. Kalaivani K, Ramachandran P. Time trends in the prevalence of anaemia in pregnancy. *Indian J Med Res.* 2018;147(3):268–277.
7. Kapil U, Kapil R, Gupta A. National Iron Plus Initiative: Current status and future strategy. *Indian J Med Res.* 2019;150(3):239–247.
8. Kassebaum NJ, Jasrasaria R, Naghavi M, et al. A systematic analysis of global anaemia burden. *Blood.* 2014;123(5):615–624.https://doi.org/10.1182/blood-2013-06-508325.
9. Let S, Tiwari S, Singh A, Chakrabarty M. Determinants of anaemia among women of reproductive age in Aspirational Districts of India. *BMC Public Health.* 2024;24:112.
10. Mog M, Neogi D, Bharadwaz MP, Panda BK, Sil A. Anaemia among married women in Northeast India. *J Biosoc Sci.* 2023;55(2):188–204.
11. Molloy AM, Kirke PN, Brody LC, Scott JM, Mills JL. Folate and vitamin B12 deficiencies and pregnancy outcomes. *Food Nutr Bull.* 2008;29(2 Suppl):S101–S111. https://doi.org/10.1177/15648265080292S114.
12. Patel R, Gupta A, Chauhan S, Bansod DW. Sanitation practices and adverse pregnancy outcomes in India. *BMC Pregnancy Childbirth.* 2019;19:430. https://doi.org/10.1186/s12884-019-2589-x PMid:31796067.
13. Sehgal R, et al. Intestinal parasitic infections and anaemia in low-income settings. *Rev Infect Dis.* 2010;32(1):e1–e7.
14. Suryanarayana R, Santhuram AN, Chandrappa M, Shivajirao P, Rangappa SS. Anaemia among pregnant women in rural Kolar. *Int J Med Sci Public Health.* 2016;5(3):454–458.
15. WHO. *Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity.* Geneva: World Health Organisation; 2011.