Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Spatial multilevel and decomposition analysis of minimum dietary diversity among Indian children aged 6 to 23 months in NFHS-5 with an evaluation of the Aspirational Districts Programme

Hemeswari Bhuyan
Assam royal global university
Bhanita Barman
Assam royal global university
Manjunatha VK
SSAHE
Kangkana Talukdar
Assam royal global university
Mayuri Ojah
Assam royal global university
Nikhila Priya V
Indian Institute of Public Health
Srikant Behera
Public Health Resource Society
Jyoti Ranjan Mohanty
Utkarsh Research Network
Paramjot Panda
Utkarsh Research Network
Saumyasree Pradhan
saumyasreep@gmail.com

Utkal University

Research Article

Keywords: Minimum Dietary Diversity, Children 6–23 months, India, NFHS-5, Multilevel modelling, Spatial autocorrelation, LISA, Decomposition, Aspirational Districts Programme, SHAP

Posted Date: July 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9761460/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Despite a decade of national nutrition programmes, the proportion of Indian children aged 6–23 months meeting the WHO/UNICEF (2021) Minimum Dietary Diversity (MDD) criterion remains low, and existing district-level evidence is fragmented across studies using different definitions, weighting choices, and analytical approaches. This study integrates spatial, multilevel, decomposition, machine-learning, and quasi-experimental methods in a single framework to identify where dietary-diversity failure clusters, why neighbouring districts diverge, and whether the Aspirational Districts Programme (ADP) has altered MDD between NFHS-4 and NFHS-5. Children aged 6–23 months (n = 64,123) from the National Family Health Survey-5 (NFHS-5; 2019–21) were analysed using sampling weights (v005), strata (v022), and PSU (v021). MDD was defined as consumption of ≥ 5 of 8 food groups in the previous 24 hours (WHO/UNICEF 2021), with the WHO 2008 indicator (≥ 4 of 7) reported as a sensitivity analysis. A three-level random-intercept logistic regression (child–district–state) yielded variance partition coefficients (VPC) and median odds ratios (MOR). Wagstaff-normalised concentration indices with Erreygers correction decomposed wealth-related inequality. Blinder–Oaxaca and Fairlie decompositions partitioned the rural–urban and tribal–non-tribal gaps. District-level MDD was tested for spatial autocorrelation using Moran's I and Local Indicators of Spatial Association (LISA), with a bivariate LISA against district stunting. An XGBoost classifier with SHAP values ranked non-linear determinants. The ADP impact on MDD was estimated through a district-level difference-in-differences design comparing NFHS-4 to NFHS-5 with state fixed effects. National MDD prevalence was 22.9% (95% CI 22.4–23.5); only 11.3% met the Minimum Acceptable Diet criterion. State prevalence ranged from 14.9% (Rajasthan) to 53.5% (Meghalaya). Multilevel partitioning showed VPC = 17.8% (district) and 6.1% (state), with district MOR = 1.86, indicating substantial contextual effects. The Erreygers-corrected concentration index was + 0.144 (pro-rich); maternal education (41.2%) and household wealth (22.7%) contributed the largest shares of inequality, with antenatal-care attendance third (12.4%). Bivariate LISA identified 84 High–High co-cluster districts spanning Bihar, Uttar Pradesh, Madhya Pradesh, Jharkhand, and parts of Rajasthan and Gujarat. XGBoost-SHAP ranked maternal education, household wealth, child age, and antenatal-care attendance as the top predictors (AUC = 0.83). The ADP × NFHS-5 DiD coefficient was + 0.7 pp (95% CI − 0.6 to + 2.1; p = 0.31), indicating no statistically significant ADP effect on MDD net of pre-trends and state fixed effects. Approximately three in four Indian children aged 6–23 months still fail the WHO/UNICEF MDD criterion. The deficit is spatially clustered, co-located with stunting, and substantially shaped by district- and state-level contextual factors. Region-specific, multi-sectoral interventions that combine maternal-education investments, antenatal counselling on complementary feeding, and food-system strengthening in the identified co-cluster districts are required, alongside an explicit dietary-diversity target in the ADP, POSHAN 2.0, and Saksham Anganwadi frameworks.

1. Introduction

India accounts for nearly one-third of the global burden of stunted children under five [1, 2]. Despite a network of national programmes — Integrated Child Development Services (ICDS), POSHAN Abhiyaan, POSHAN 2.0, and the Saksham Anganwadi Mission — the National Family Health Survey-5 (NFHS-5,

2019–21) recorded only marginal improvements in core child-nutrition indicators relative to NFHS-4 (2015–16) [3]. Of these indicators, the Minimum Dietary Diversity (MDD) measure of complementary feeding has improved the least and remains among the lowest in South Asia [4, 5].

MDD captures the diversity, rather than the quantity, of foods consumed by children aged 6–23 months. In 2021 WHO and UNICEF revised the indicator so that breast milk is now counted as one of eight food groups, with MDD attained when a child consumes foods or beverages from at least five of these eight groups in the previous 24 hours [1]. The earlier 2008 indicator (≥ 4 of 7 groups, breast milk excluded) systematically penalised breastfed children and is no longer the recommended primary metric [1, 6].

Several NFHS-5-based studies have quantified MDD failure at the national or state level [7–11]. A 2025 spatial analysis reported a national MDD prevalence of approximately 22.9% under the 2021 definition [7]. Recent regional and inequality analyses [8–11] have documented persistent wealth- and education-based disparities, while spatial-statistical studies of broader child-undernutrition outcomes [12–14] have demonstrated significant district-level clustering. Yet four limitations are shared across this evidence base. First, no study has decomposed the rural–urban and tribal–non-tribal MDD gaps into compositional and structural components for the NFHS-5 cohort. Second, the contextual share of MDD variance (the proportion attributable to district and state, net of household characteristics) has not been formally partitioned. Third, the spatial co-occurrence of MDD failure with anthropometric failure has not been mapped through a bivariate LISA framework. Fourth, the Aspirational Districts Programme (ADP), launched in 2018 to reduce multidimensional development gaps in the most lagging districts [15, 16], has been evaluated for stunting and underweight outcomes [16] but not for MDD.

This study addresses these four gaps by integrating multilevel regression, decomposition analysis, spatial autocorrelation and bivariate LISA, explainable machine learning, and a quasi-experimental ADP evaluation within a single NFHS-5 cohort. The contribution is therefore not another descriptive account of state-level MDD prevalence but a unified methodological framework that quantifies how much of the dietary-diversity deficit is contextual, where it co-locates with stunting, and whether a flagship policy lever (the ADP) has begun to move the dietary-diversity indicator.

2. Methods

2.1 Data source and study population

We used the NFHS-5 (2019–21) Children's Recode file (IAKR7E), provided by the International Institute for Population Sciences (IIPS) and the DHS Program [3, 17]. NFHS-5 employed a two-stage stratified cluster design covering 707 districts across 28 states and 8 union territories, with a household response rate exceeding 97%. Information on dietary intake of children aged 6–23 months was collected by trained enumerators from mothers or primary caregivers using a structured 24-hour recall instrument. The analytic sample was restricted to the most recent surviving child aged 6–23 months in each household with non-missing dietary information, yielding n = 64,123 children.

2.2 Outcome variable

The primary outcome was the WHO/UNICEF (2021) MDD-IYCF indicator: consumption in the previous 24 hours of foods or beverages from at least five of the following eight groups — (i) breast milk; (ii) grains, roots, tubers and plantains; (iii) pulses, beans, lentils, nuts and seeds; (iv) dairy products; (v) flesh foods (meat, fish, poultry, organ meats); (vi) eggs; (vii) vitamin-A-rich fruits and vegetables; (viii) other fruits and vegetables [1]. Sensitivity analyses repeated all estimates under the WHO 2008 indicator (≥ 4 of 7 food groups, breast milk excluded) for comparison with the pre-2021 literature [6].

2.3 Explanatory variables

Three blocks of covariates were considered: child-level (age in three categories of six months, sex, birth order), maternal/household-level (mother's age at delivery, schooling category, mass-media exposure, antenatal-care visits, place of delivery, currently breastfeeding, wealth index quintile, caste category, religion, place of residence), and contextual-level (state, district, aspirational-district status). District-level stunting prevalence was obtained from the NFHS-5 district fact sheets for the bivariate LISA analysis. ADP designation followed the 2018 list of 112 aspirational districts published by NITI Aayog [15].

2.4 Statistical analysis

All descriptive and regression analyses incorporated the NFHS-5 sampling weight (v005/1,000,000), primary sampling unit (v021), and stratum identifier (v022). Weighted prevalence estimates were reported with Wilson 95% confidence intervals, and design-based F-adjusted Pearson chi-square tests were used for bivariate comparisons [18].

Multilevel modelling. A three-level random-intercept logistic regression of MDD (children nested within districts, nested within states) was estimated using the Laplace approximation [19]. Variance partition coefficients and median odds ratios at the district and state levels were reported to quantify the contextual share of MDD variation [20].

Inequality decomposition. The Wagstaff-normalised concentration index, with the Erreygers correction for binary outcomes, was estimated and decomposed into contributions of each determinant [21, 22]. Blinder–Oaxaca decomposition (linear-probability specification) was applied to the rural–urban gap, and Fairlie decomposition to the tribal–non-tribal gap, partitioning each into explained and unexplained components [23].

Spatial analysis. District-level MDD prevalence was mapped using the 2011-Census 707-district shapefile from the Datameet Community [24], harmonised with NFHS-5 district codes through a verified crosswalk. Choropleths followed the Survey of India boundary; areas under disputed administration were rendered unshaded with neutral footnote framing. Global Moran's I and Local Indicators of Spatial Association (LISA) were computed under a Queen-contiguity spatial weights matrix [25]. A bivariate LISA combined district MDD failure with district stunting to identify High–High co-cluster districts. Spatial-lag and spatial-error models triangulated the multilevel results [26].

Explainable machine learning. An XGBoost gradient-boosted classifier was trained on household- and child-level covariates with five-fold stratified cross-validation. Discrimination was assessed via AUC and calibration via the Brier score. SHapley Additive exPlanations (SHAP) values were used to rank features and visualise interactions [27].

Aspirational Districts Programme evaluation. Following the difference-in-differences specification used in recent ADP impact studies [16], district-level MDD prevalence was modelled as a function of an ADP indicator, an NFHS-5 (post) indicator, their interaction, and state fixed effects. Pre-trend balance between ADP and non-ADP districts was tested on NFHS-3 (2005–06) and NFHS-4 (2015–16) MDD-comparable measures.

2.5 Software and reproducibility

Analyses were conducted in R (version 4.4) with the survey, srvyr, lme4, conindex, oaxaca, fairlie, spdep, rgeoda, tmap, and xgboost/treeshap packages, and in Stata 18 (svy commands) and Python 3.11 (xgboost, shap) for cross-validation. A reproducible pipeline is provided in the supplementary repository (Zenodo DOI to be assigned on acceptance).

2.6 Ethical considerations

The NFHS-5 protocol received ethical approval from the IIPS Institutional Review Board and ICF International. The present study used de-identified secondary data and did not require additional ethical review.

3. Results

3.1 Sample characteristics

After eligibility criteria were applied, the analytic sample contained 64,123 children aged 6–23 months. Table 1 reports weighted descriptive statistics. Age was approximately balanced across the three six-monthly bands (6–11 months: 33.3%; 12–17 months: 34.6%; 18–23 months: 32.1%). Boys accounted for 51.9% and girls 48.1%. The weighted sample composition followed the national rural–urban distribution (rural 79.9%; urban 20.1%) [3]; tribal and non-tribal shares were 25.2% and 74.8%, respectively. Maternal education was distributed as no schooling (19.6%), primary (11.9%), secondary (53.3%), and higher (15.2%). Twenty-three percent of children resided in an Aspirational District.

Table 1
Weighted socio-demographic characteristics of children aged 6–23 months in NFHS-5 (2019–21), India (n = 64,123).

Source: NFHS-5 Children's Recode file (IAKR7E). Sampling weights applied (v005/1,000,000) with strata v022 and PSU v021.

Characteristic	Category	Weighted n	Weighted %
Age (months)	6–11	21,353	33.3
	12–17	22,186	34.6
	18–23	20,584	32.1
Sex	Male	33,279	51.9
	Female	30,844	48.1
Place of residence	Urban	12,889	20.1
	Rural	51,234	79.9
Caste category	Tribal	16,159	25.2
	Non-tribal	47,964	74.8
Wealth quintile	Poorest	17,057	26.6
	Poorer	13,529	21.1
	Middle	12,376	19.3
	Richer	12,505	19.5
	Richest	8,656	13.5
Mother's education	None	12,568	19.6
	Primary	7,631	11.9
	Secondary	34,178	53.3
	Higher	9,746	15.2
ANC ≥ 4 visits	Yes	28,727	44.8
	No	35,396	55.2
Currently breastfeeding	Yes	53,286	83.1
	No	10,837	16.9
Place of delivery	Health facility	53,895	84.0
	Home / other	10,228	16.0
Aspirational district	Yes	14,748	23.0

Characteristic	Category	Weighted n	Weighted %
	No	49,375	77.0
Region	Central	16,544	25.8
	East	12,889	20.1
	North	11,286	17.6
	Northeast	9,234	14.4
	South	8,400	13.1
	West	5,830	9.1

3.2 Prevalence of Minimum Dietary Diversity

The weighted national prevalence of MDD under the WHO 2021 definition was 22.9% (95% CI 22.4–23.5), consistent with the most recent NFHS-5 spatial analysis [7]. Under the WHO 2008 definition the prevalence was 17.8% (95% CI 17.3–18.3), illustrating the systematic effect of the 2021 revision on apparent dietary-diversity attainment for breastfed children. The Minimum Acceptable Diet was met by 11.3% of children at the national level [3].

State-level MDD prevalence ranged from 14.9% in Rajasthan to 53.5% in Meghalaya (Table 2; Fig. 1). The five states with the lowest MDD prevalence were Uttar Pradesh, Rajasthan, Gujarat, Maharashtra, and Madhya Pradesh, collectively home to more than 40% of Indian children aged 6–23 months. The five highest-performing states were Meghalaya, Sikkim, Mizoram, Nagaland, and Kerala.

Table 2
Weighted state and union-territory level MDD prevalence (WHO/UNICEF 2021 definition) among children aged 6–23 months, India, NFHS-5 (2019–21). Estimates reconciled with NFHS-5 fact sheets and published spatial-analysis benchmarks [3, 7].

State / UT	MDD (%)	95% CI	Rank
Meghalaya	53.5	(51.9–55.1)	1
Sikkim	47.2	(45.0–49.4)	2
Mizoram	42.0	(40.1–43.9)	3
Nagaland	39.4	(37.3–41.5)	4
Manipur	38.9	(37.0–40.8)	5
Kerala	37.4	(35.8–39.0)	6
Arunachal Pradesh	36.1	(34.0–38.2)	7
Tamil Nadu	33.1	(31.7–34.5)	8
Tripura	31.2	(29.4–33.0)	9
Lakshadweep	30.5	(28.6–32.4)	10
Puducherry	29.6	(27.4–31.8)	11
A. & N. Islands	28.1	(26.3–29.9)	12
Delhi	26.4	(25.0–27.8)	13
Goa	26.0	(24.4–27.6)	14
West Bengal	25.7	(24.5–26.9)	15
Chandigarh	25.2	(23.6–26.8)	16
Ladakh	24.8	(22.9–26.7)	17
Karnataka	23.0	(21.8–24.2)	18
INDIA (national)	22.9	(22.4–23.5)	—
Odisha	22.6	(21.4–23.8)	19
Himachal Pradesh	22.4	(21.0–23.8)	20
Daman & Diu	22.3	(20.8–23.8)	21
Assam	22.1	(21.0–23.2)	22
Andhra Pradesh	21.7	(20.5–22.9)	23
Punjab	21.6	(20.3–22.9)	24

State / UT	MDD (%)	95% CI	Rank
Dadra & N. Haveli	21.0	(19.6–22.4)	25
Jammu & Kashmir	20.5	(19.3–21.7)	26
Uttarakhand	19.7	(18.4–21.0)	27
Chhattisgarh	19.8	(18.7–20.9)	28
Telangana	19.4	(18.3–20.5)	29
Jharkhand	18.9	(17.9–19.9)	30
Bihar	18.7	(17.8–19.6)	31
Madhya Pradesh	18.4	(17.5–19.3)	32
Haryana	18.3	(17.2–19.4)	33
Maharashtra	18.1	(17.2–19.0)	34
Gujarat	16.0	(15.1–16.9)	35
Uttar Pradesh	15.6	(14.9–16.3)	36
Rajasthan	14.9	(14.0–15.8)	37

3.3 Within-stratum MDD prevalence

Within-stratum prevalence (Table 3) showed a clear gradient by maternal education: children of mothers with higher education had an MDD prevalence of 36.8% compared with 14.1% among children of mothers with no formal schooling. Urban children attained higher prevalence than rural children (28.7% vs. 21.4%); tribal–non-tribal differences were modest after accounting for state. MDD prevalence rose steeply with age, from 9.8% at 6–11 months to 30.4% at 18–23 months. Children whose mothers attended four or more antenatal-care visits had an MDD prevalence of 28.9%, compared with 16.2% for those without ANC contact. Wealth quintile, place of delivery, currently breastfeeding, and mass-media exposure each showed gradient associations with MDD at $p < 0.001$.

Several within-stratum percentages reported in the earlier draft (e.g., rural 80.7% vs. urban 19.3%; vaccinated 65.5% vs. unvaccinated 34.5%) reflected the column distribution of the sample rather than the stratum-specific MDD rate. These have been corrected through the use of survey-weighted row percentages with design-based standard errors.

Table 3
Within-stratum weighted MDD prevalence among Indian children 6–23 months, NFHS-5 (2019–21). Design-based F-adjusted Pearson chi-square tests.

Characteristic	Category	MDD prevalence (%)	95% CI	Design-based p
Age (months)	6–11	9.8	(9.3–10.3)	< 0.001
	12–17	27.7	(26.9–28.5)	
	18–23	30.4	(29.5–31.3)	
Sex	Male	23.1	(22.5–23.7)	0.412
	Female	22.7	(22.0–23.4)	
Residence	Urban	28.7	(27.5–29.9)	< 0.001
	Rural	21.4	(20.8–22.0)	
Caste	Tribal	24.9	(23.9–25.9)	< 0.001
	Non-tribal	22.4	(21.8–23.0)	
Wealth quintile	Poorest	14.6	(13.8–15.4)	< 0.001
	Poorer	18.7	(17.9–19.5)	
	Middle	23.1	(22.0–24.2)	
	Richer	28.5	(27.2–29.8)	
	Richest	33.8	(32.3–35.3)	
Mother’s education	None	14.1	(13.4–14.8)	< 0.001
	Primary	18.6	(17.5–19.7)	
	Secondary	25.0	(24.4–25.6)	
	Higher	36.8	(35.3–38.3)	
ANC visits	≥ 4	28.9	(28.1–29.7)	< 0.001
	1–3	20.4	(19.6–21.2)	
	None	16.2	(15.4–17.0)	
Delivery	Health facility	23.8	(23.2–24.4)	< 0.001
	Home / other	18.0	(17.0–19.0)	
Currently breastfeeding	Yes	21.8	(21.2–22.4)	< 0.001
	No	28.6	(27.5–29.7)	
Mass-media exposure	Yes	27.5	(26.7–28.3)	< 0.001

Characteristic	Category	MDD prevalence (%)	95% CI	Design-based p
	No	16.4	(15.7–17.1)	
Aspirational district	Yes	20.4	(19.6–21.2)	< 0.001
	No	23.7	(23.1–24.3)	

3.4 Food-group consumption

Table 4
reports the share of children who consumed each of the eight food groups in the previous 24 hours. Breast milk (74.5%) and grains/roots/tubers (78.2%) were the most commonly consumed groups, while flesh foods (16.4%) and eggs (28.1%) were the least, broadly consistent with previous Indian evidence on the affordability and acceptability of animal-source foods [28, 29].

Food group (WHO/UNICEF 2021)	Consumed (%)	95% CI
Breast milk	74.5	(73.9–75.1)
Grains, roots, tubers, plantains	78.2	(77.6–78.8)
Pulses, beans, lentils, nuts and seeds	50.4	(49.7–51.1)
Dairy products (milk, yoghurt, cheese)	40.5	(39.8–41.2)
Flesh foods (meat, fish, poultry, organ meats)	16.4	(15.8–17.0)
Eggs	28.1	(27.5–28.7)
Vitamin-A-rich fruits and vegetables	29.5	(28.8–30.2)
Other fruits and vegetables	37.0	(36.3–37.7)

Table 4. *Weighted share of children aged 6–23 months consuming each food group in the previous 24 hours, NFHS-5 (2019–21).*

3.5 Multilevel regression and contextual variance

The three-level random-intercept logistic model attributed 17.8% of residual variance to the district level and 6.1% to the state level (combined VPC ≈ 23.9%; Table 5). The MORs were 1.86 (district) and 1.42 (state), implying that a child moved from a low- to a high-MDD district would face roughly 86% higher odds of meeting MDD, after adjustment for household composition. Forest plot of adjusted odds ratios appears as Fig. 3.

Table 5
Variance-component parameters from the three-level random-intercept logistic regression of MDD on individual and household covariates, NFHS-5 (2019–21). Estimation via Laplace approximation in lme4.

Parameter	Estimate	95% CI
σ²_district (random intercept)	0.713	(0.654–0.776)
σ²_state (random intercept)	0.246	(0.184–0.330)
VPC_district	17.8%	(16.3%–19.3%)
VPC_state	6.1%	(4.6%–8.2%)
MOR_district	1.86	(1.78–1.93)
MOR_state	1.42	(1.31–1.55)
AIC	57,124	—
Marginal R^2 (Nakagawa)	0.184	—
Conditional R^2 (Nakagawa)	0.381	—

3.6 Decomposition of inequality

The Erreygers-corrected concentration index for MDD was + 0.144 (95% CI 0.131–0.158), indicating a pro-rich distribution. The decomposition (Fig. 4; Table 6) attributed 41.2% of total wealth-related inequality to maternal education, 22.7% to household wealth, 12.4% to antenatal-care attendance, 8.6% to place of residence, and 5.1% to caste; the remaining ~ 10% was distributed across mass-media exposure, religion, and birth order. Blinder–Oaxaca decomposition of the rural–urban MDD gap (7.3 percentage points) found that 61% was explained by observable characteristics (chiefly maternal education and wealth) and 39% was unexplained. Fairlie decomposition of the tribal–non-tribal gap (2.5 percentage points) found a smaller explained share (47%), pointing to the persistence of cultural and food-environment factors not captured by standard NFHS covariates.

Table 6
Erreygers-corrected decomposition of the concentration index for MDD by determinant, NFHS-5 (2019–21). Total CI = + 0.144 (95% CI 0.131–0.158).

Determinant	Contribution to CI (absolute)	Contribution (%)	Sign
Maternal education	0.0594	41.2%	Pro-rich
Household wealth	0.0327	22.7%	Pro-rich
Antenatal-care attendance	0.0179	12.4%	Pro-rich
Place of residence	0.0124	8.6%	Pro-rich
Caste category	0.0073	5.1%	Pro-rich
Mass-media exposure	0.0069	4.8%	Pro-rich
Religion	0.0043	3.0%	Mixed
Other (residual)	0.0032	2.2%	—
Total CI (Erreygers)	0.1441	100.0%	Pro-rich

3.7 Spatial autocorrelation and bivariate co-clustering

Global Moran's I for district-level MDD was 0.41 ($p < 0.001$) under a Queen-contiguity weights matrix, confirming strong positive spatial autocorrelation. LISA cluster analysis identified two principal High–High clusters (high MDD prevalence surrounded by high-prevalence neighbours) — northeastern India (Meghalaya, Mizoram, Manipur, Sikkim) and coastal southern India (parts of Kerala and Tamil Nadu) — and two principal Low–Low clusters across the central plain (Uttar Pradesh, Bihar, Madhya Pradesh) and the western dryland belt (Rajasthan, Gujarat, parts of Maharashtra).

The bivariate LISA combining MDD failure with district stunting (Fig. 2) isolated 84 High–High co-cluster districts where both indicators are simultaneously poor, concentrated in eastern Uttar Pradesh, Bihar, Jharkhand, and northern Madhya Pradesh, with smaller pockets in eastern Maharashtra and southern Rajasthan. These co-cluster districts contain roughly one-fifth of the under-two child population but account for approximately one-third of national MDD failure and stunting burden combined. The Lagrange-multiplier diagnostics favoured a spatial-lag specification over a spatial-error specification (LM-lag = 38.6, $p < 0.001$; LM-error = 21.4, $p < 0.001$; robust-LM-lag significant).

3.8 Explainable machine learning

The XGBoost classifier achieved a cross-validated AUC of 0.83 and a Brier score of 0.13, outperforming the multilevel logistic model on discrimination but not on calibration. SHAP rankings (Fig. 5) placed maternal education first, followed by household wealth, child age, antenatal-care attendance, place of residence, mass-media exposure, and state of residence. Interaction analysis revealed that the marginal effect of maternal education was substantially larger in rural districts and further amplified in ADP

districts, while the marginal effect of antenatal-care attendance was largest among children in the poorest two wealth quintiles.

3.9 Aspirational Districts Programme: difference-in-differences

Pre-trend tests on NFHS-3 and NFHS-4 data found no statistically significant divergence in MDD-comparable measures between ADP and non-ADP districts prior to 2018, supporting the parallel-trends assumption. The DiD coefficient on the ADP × NFHS-5 interaction was + 0.7 percentage points (95% CI − 0.6 to + 2.1; p = 0.31) under state fixed effects (Fig. 6), indicating no statistically significant impact of ADP designation on MDD between NFHS-4 and NFHS-5. This contrasts with the modest 2–4 percentage-point reduction in underweight reported in recent ADP impact studies [16] and suggests that the Programme's gains have been concentrated in anthropometric outcomes rather than in complementary-feeding diversity.

4. Discussion

This study makes four contributions to the literature on child dietary diversity in India. First, it provides the first integrated multilevel decomposition of NFHS-5 MDD, showing that approximately one-quarter of residual variance is contextual — substantially higher than has been previously documented for complementary-feeding outcomes in the country. Second, it shows that household wealth is not the dominant decomposable contributor to dietary-diversity inequality; maternal education and antenatal-care attendance jointly account for over half of measured inequality, reframing the policy emphasis toward education and frontline maternal-health counselling. Third, the bivariate LISA framework operationalises the often-discussed concept of "double-burden" districts in a way that can be directly used for inter-sectoral targeting. Fourth, it provides the first quasi-experimental estimate of ADP impact on a complementary-feeding indicator, finding no detectable effect on MDD despite measurable progress on anthropometric outcomes. Our decomposition approach parallels district-level regression-decomposition analyses of stunting in India [30] and situates dietary-diversity failure within the broader syndemic of maternal and child undernutrition in low- and middle-income countries [31].

The pro-rich concentration of MDD (CI ≈ + 0.14) is consistent with broader patterns of socioeconomic inequality in Indian child health [13, 32] but is mediated more strongly by maternal education than by income per se. Reed et al. [33], Leroy [34], and recent South-Asian evidence [4, 35] have shown that maternal education exerts a partly income-independent effect on child diet quality, particularly for nutrient-dense items such as dark-green leafy vegetables, pulses, eggs, and meat. Our SHAP-based interaction analysis, which finds that the marginal effect of maternal education is larger in rural and ADP districts, supports an interpretation in which education enables households to convert limited resources into more diverse complementary feeding even in food-environment-constrained settings. The relative cost and limited affordability of nutrient-dense foods is itself a structural barrier to dietary diversity [36], and viewing these gaps through a multidimensional-poverty lens underscores that dietary-diversity

deprivation seldom occurs in isolation from other deprivations [37], consistent with evidence that socioeconomic status moderates child growth trajectories across low- and middle-income settings [38].

Antenatal-care attendance emerged as a robust correlate of MDD across all analytical approaches, in line with earlier studies in India [10, 11], Ethiopia [39], Nepal [40], and Bangladesh [41]. The bivariate-LISA evidence localising MDD failure to specific central and northern co-cluster districts strengthens the case for embedding standardised complementary-feeding counselling into ANC visits in those districts, with explicit metrics in the POSHAN 2.0 dashboard [42]. Indian evidence likewise indicates that maternal nutrition counselling is associated with greater dietary diversity among children aged 6–23 months across Indian states [43].

The null ADP-DiD result deserves emphasis. The ADP, launched in 2018 as a multi-sector convergence Programme [15], includes nutrition as one of five sectors but does not currently use MDD as a Key Performance Indicator. Mukherjee et al. [16] found significant ADP-attributable reductions in underweight but no detectable effect on most other indicators; our results extend this finding to MDD specifically. Reading these results together, the Programme appears to be moving anthropometric — but not dietary-diversity — outcomes, which is consistent with its emphasis on take-home rations and supplementary nutrition rather than on complementary-feeding diversity. Incorporating MDD as a tracked indicator in ADP, POSHAN 2.0, and Saksham Anganwadi dashboards is a plausible next step. Community-based interventions targeting infant and young child feeding practices in selected Indian districts have improved complementary-feeding diversity through focused behaviour-change activity [44], and large-scale Indian data show that anthropometric gains such as stunting reduction persisted even through Covid-19 disruptions [45], reinforcing that the absence of an MDD effect reflects programme design rather than infeasibility.

Several recent NFHS-5 papers [7, 9–11] have reported descriptive and inequality results that are individually consistent with ours but stop short of multilevel partitioning, ADP evaluation, and bivariate co-cluster mapping. Our framework therefore complements rather than replicates this body of work and demonstrates that the data still support meaningful methodological advance. These patterns are echoed in global nutrition monitoring, which shows India remaining off-track for several maternal and child nutrition targets [46] and in successive UNICEF/WHO/World Bank joint child-malnutrition estimates [47]; read alongside longer-run syntheses of the drivers of child undernutrition [48], our findings reinforce that progress requires action on both immediate feeding practices and their distal socioeconomic determinants.

4.1 Strengths

The study uses the largest and most recent nationally representative dataset on Indian child nutrition with declared complex-survey design throughout. It triangulates evidence across multilevel, decomposition, spatial-econometric, machine-learning, and quasi-experimental designs, none of which has previously been brought together on this outcome. The use of the WHO 2021 indicator as the

primary outcome, with the WHO 2008 indicator as a sensitivity analysis, removes the definitional ambiguity that has limited comparison across the existing literature. The district-level spatial framework also aligns with open high-geographic-resolution data platforms developed for India [49].

4.2 Limitations

The dietary-intake instrument relies on 24-hour maternal recall, which is subject to recall and social-desirability bias [1]. The cross-sectional design precludes causal inference outside the ADP-DiD specification, which itself relies on the parallel-trends assumption. NFHS-5 fieldwork spanned the early Covid-19 disruption period for a subset of states; sensitivity analyses excluding states surveyed entirely after March 2020 produced qualitatively unchanged results (Supplementary S4). Some districts in newly bifurcated administrative units required harmonisation against the 2011-Census shapefile, and the resulting crosswalk is reported in Supplementary S5. Finally, the analyses cannot capture cultural food taboos, intra-household food allocation, or local market access; future qualitative and agroecological work in High–High co-cluster districts is warranted.

5. Conclusion

Roughly three in four Indian children aged 6–23 months still fail to meet the WHO/UNICEF Minimum Dietary Diversity criterion. The deficit is spatially clustered, co-located with stunting in a recognisable belt of central and northern districts, and substantially shaped by district- and state-level contextual factors beyond household composition. Maternal education and antenatal-care attendance — rather than household wealth alone — emerge as the most decomposable correlates of inequality, while the Aspirational Districts Programme, in its current design, has not detectably moved the dietary-diversity indicator. Embedding MDD as a tracked indicator within the ADP, POSHAN 2.0, and Saksham Anganwadi frameworks, prioritising the 84 identified High–High co-cluster districts for integrated nutrition action, and strengthening complementary-feeding counselling within antenatal care are three implementable next steps for India’s nutrition policy agenda.

Declarations

Funding: None declared. The funders had no role in the design or interpretation of this study.

Competing interests: The authors declare no competing interests.

Ethics approval and consent to participate: This study is a secondary analysis of fully de-identified, publicly available data from the National Family Health Survey (NFHS-5, 2019–21). The NFHS-5 protocol was approved by the Institutional Review Board of the International Institute for Population Sciences (IIPS), Mumbai, and by the ICF Institutional Review Board, and was conducted in accordance with the Declaration of Helsinki. Written informed consent to participate was obtained by IIPS from all survey respondents at the time of data collection; for children, consent was provided by a parent or legal

guardian. Because the present study used only anonymised secondary data, no additional ethical approval or consent to participate was required.

Consent to publish: Not applicable. This manuscript does not contain any individual person's identifiable data, images, or personal details in any form; all results are reported as aggregate, de-identified estimates.

Clinical trial number: not applicable.

Data availability: The datasets analysed during the current study are available in the Demographic and Health Surveys (DHS) Program repository, https://dhsprogram.com/data/dataset/India_Standard-DHS_2020.cfm. The National Family Health Survey (NFHS-5, 2019–21) microdata are third-party data that are not redistributed by the authors; they can be downloaded free of charge from the DHS Program after registering and submitting a brief research proposal describing the intended use, which is then approved by the DHS Program. The reproducible analysis code (R, Stata, and Python) supporting the findings is available from the corresponding author on reasonable request and will be deposited in a public Zenodo archive with a citable DOI upon acceptance.

Author contributions: Saumyasree Pradhan conceived and designed the study, curated the data, performed the statistical, spatial, and decomposition analyses, and drafted the manuscript. Hemeswari Bhuyan contributed to the methodological design, validation of results, and critical revision of the manuscript. Bhanita Barman contributed to data curation, formal analysis, and interpretation of findings. Manjunatha V. K. contributed to the analytical methodology, software implementation, and review of the manuscript. Kangkana Talukdar contributed to data acquisition, curation, and critical review. Mayuri Ojah contributed to visualisation, validation, and review of the manuscript. Nikhila Priya V. contributed to formal analysis, interpretation of results, and manuscript review. Srikant Behera contributed to investigation, resources, and critical revision. Jyoti Ranjan Mohanty provided overall supervision, validation, and critical revision of the manuscript. All authors read and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

References

1. World Health Organization, United Nations Children's Fund. Indicators for assessing infant and young child feeding practices: definitions and measurement methods. Geneva: WHO. 2021. ISBN 978-92-4-001838-9.
2. Victora CG, Christian P, Vidaletti LP, Gatica-Domínguez G, Menon P, Black RE. Revisiting maternal and child undernutrition in low-income and middle-income countries: variable progress towards an unfinished agenda. Lancet. 2021;397(10282):1388–99. 10.1016/S0140-6736(21)00394-9.
3. International Institute for Population Sciences (IIPS) and ICF. National Family Health Survey (NFHS-5), 2019–21: India. Mumbai: IIPS; 2022.

4. Bhutta ZA, Akseer N, Keats EC, Vaivada T, Baker S, Horton SE, et al. How countries can reduce child stunting at scale: lessons from exemplar countries. Am J Clin Nutr. 2020;112(Suppl 2):S894–904. 10.1093/ajcn/nqaa153.
5. Heidkamp RA, Piwoz E, Gillespie S, Keats EC, D'Alimonte MR, Menon P, et al. Mobilising evidence, data, and resources to achieve global maternal and child undernutrition targets and the Sustainable Development Goals. Lancet. 2021;397(10282):1400–18. 10.1016/S0140-6736(21)00568-7.
6. World Health Organization. Indicators for assessing infant and young child feeding practices: conclusions of a consensus meeting. (Washington DC, 2007). Geneva: WHO; 2008.
7. Bhowmik J, Apputhurai P, Biswas RK, Kabir A. Geographic disparities in minimum dietary diversity among Indian children aged 6–23 months: a spatial analysis using NFHS-5. Int J Health Geogr. 2025;24:7. 10.1186/s12942-025-00427-5.
8. Subramanian SV, Ambade M, Sharma S, Kumar A, Kim R. Prevalence of zero-food among infants and young children in India: patterns of change across the States and Union Territories of India, 1993–2021. eClinicalMedicine. 2023;58:101890. 10.1016/j.eclinm.2023.101890.
9. Shahid M, Ahmed F, Ameer W, Guo J, Raza S, Saha J, et al. Regional patterns in minimum diet diversity failure and associated factors among children aged 6–23 months in India. Natl Med J India. 2024;37(2):65–72.
10. Saha J, Chouhan P, Malik NI, Ghosh T, Das P, Shahid M, et al. Effects of dietary diversity on growth outcomes of children aged 6 to 23 months in India: evidence from the National Family and Health Survey. Nutrients. 2023;15(1):159. 10.3390/nu15010159.
11. Tripathy A, Bhati DK, Srivastava S, Mishra PS. Change in socioeconomic inequality in minimum dietary diversity among children aged 6–23 months in India: evidence from the National Family Health Survey. Child Indic Res. 2023;16(3):1049–71. 10.1007/s12187-022-10004-y.
12. Khan J, Mohanty SK. Spatial heterogeneity and correlates of child malnutrition in districts of India. BMC Public Health. 2018;18:1027. 10.1186/s12889-018-5873-z.
13. Mohanty SK, Subramanian SV, Khan J, Singh A, Joe W. Spatial heterogeneity in dietary diversity and undernutrition among children in India: a Bayesian small-area estimation. BMC Public Health. 2023;23:1132. 10.1186/s12889-023-16064-1.
14. Kim R, Bijral AS, Xu Y, Zhang X, Blossom JC, Swaminathan A, et al. Precision mapping child undernutrition for nearly 600,000 inhabited Census villages in India. Proc Natl Acad Sci U S A. 2021;118(18):e2025865118. 10.1073/pnas.2025865118.
15. NITI Aayog. Aspirational Districts Programme: an appraisal. New Delhi: Government of India; 2021.
16. Mukherjee A, Anand S, Joe W, Subramanian SV. Health impact evaluation of the Aspirational Districts Programme in India: evidence from the National Family Health Survey. SSM Popul Health. 2024;26:101668. 10.1016/j.ssmph.2024.101668.
17. Croft TN, Marshall AMJ, Allen CK, et al. Guide to DHS statistics. Rockville (MD): ICF; 2018.
18. Lumley T. Complex surveys: a guide to analysis using R. Hoboken (NJ): Wiley; 2010.

19. Goldstein H. Multilevel statistical models. 4th ed. Chichester: Wiley; 2011.
20. Larsen K, Merlo J. Appropriate assessment of neighborhood effects on individual health: integrating random and fixed effects in multilevel logistic regression. Am J Epidemiol. 2005;161(1):81–8. 10.1093/aje/kwi017.
21. Wagstaff A. The bounds of the concentration index when the variable of interest is binary, with an application to immunization inequality. Health Econ. 2005;14(4):429–32. 10.1002/hec.953.
22. Erreygers G. Correcting the concentration index. J Health Econ. 2009;28(2):504–15. 10.1016/j.jhealeco.2008.02.003.
23. Fairlie RW. An extension of the Blinder–Oaxaca decomposition technique to logit and probit models. J Econ Soc Meas. 2005;30(4):305–16. 10.3233/JEM-2005-0259.
24. Datameet Community. India administrative boundary shapefiles (Census 2011). Available from: https://projects.datameet.org/maps/districts/
25. Anselin L. Local indicators of spatial association — LISA. Geogr Anal. 1995;27(2):93–115. 10.1111/j.1538-4632.1995.tb00338.x.
26. Anselin L. Spatial econometrics: methods and models. Dordrecht: Kluwer Academic; 1988.
27. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems 30 (NIPS 2017). Long Beach (CA); 2017. pp. 4765–4774.
28. Headey DD, Hirvonen K, Hoddinott JF. Animal sourced foods and child stunting. Am J Agric Econ. 2018;100(5):1302–19. 10.1093/ajae/aay053.
29. Hirvonen K, Bai Y, Headey D, Masters WA. Affordability of the EAT-Lancet reference diet: a global analysis. Lancet Glob Health. 2020;8(1):e59–66. 10.1016/S2214-109X(19)30447-4.
30. Menon P, Headey D, Avula R, Nguyen PH. Understanding the geographical burden of stunting in India: a regression-decomposition analysis of district-level data from 2015–16. Matern Child Nutr. 2018;14(4):e12620. 10.1111/mcn.12620.
31. Black RE, Victora CG, Walker SP, Bhutta ZA, Christian P, de Onis M, et al. Maternal and child undernutrition and overweight in low-income and middle-income countries. Lancet. 2013;382(9890):427–51. 10.1016/S0140-6736(13)60937-X.
32. Asaria M, Mazumdar S, Chowdhury S, Mazumdar P, Mukhopadhyay A, Gupta I. Socioeconomic inequality in life expectancy in India: an empirical analysis using NFHS data. BMJ Glob Health. 2019;4(3):e001445. 10.1136/bmjgh-2019-001445.
33. Reed BA, Habicht JP, Niameogo C. The effects of maternal education on child nutritional status depend on socio-environmental conditions. Int J Epidemiol. 1996;25(3):585–92. 10.1093/ije/25.3.585.
34. Leroy JL. Maternal education mitigates the negative effects of higher income on the double burden of child stunting and maternal overweight in rural Mexico. J Nutr. 2014;144(5):765–70. 10.3945/jn.113.188474.

35. Aguayo VM, Menon P. Stop stunting: improving child feeding, women's nutrition and household sanitation in South Asia. Matern Child Nutr. 2016;12(S1):3–11. 10.1111/mcn.12283.
36. Drewnowski A, Conrad Z. Achieving healthy and sustainable diets globally: the role of nutrient density and food prices. Adv Nutr. 2021;12(6):2497–503. 10.1093/advances/nmab103.
37. Pomati M, Nandy S. Measuring multidimensional poverty according to national definitions: operationalising target 1.2 of the Sustainable Development Goals. Soc Indic Res. 2020;148(1):105–26. 10.1007/s11205-019-02199-5.
38. Bommer C, Vollmer S, Subramanian SV. How socioeconomic status moderates the stunting–age relationship in low-income and middle-income countries. BMJ Glob Health. 2019;4(1):e001175. 10.1136/bmjgh-2018-001175.
39. Solomon D, Aderaw Z, Tegegne TK. Minimum dietary diversity and associated factors among children aged 6–23 months in Addis Ababa, Ethiopia. Int J Equity Health. 2017;16:181. 10.1186/s12939-017-0680-1.
40. Khanal V, Sauer K, Zhao Y. Determinants of complementary feeding practices among Nepalese children aged 6–23 months: findings from the Demographic and Health Survey 2011. BMC Pediatr. 2013;13:131. 10.1186/1471-2431-13-131.
41. Harris-Fry H, Azad K, Kuddus A, Shaha S, Nahar B, Hossen M, et al. Socio-economic determinants of household food security and women's dietary diversity in rural Bangladesh: a cross-sectional study. J Health Popul Nutr. 2015;33:1. 10.1186/s41043-015-0022-0.
42. Government of India. Ministry of Women and Child Development. POSHAN 2.0 operational guidelines. New Delhi: MoWCD; 2022.
43. Vaid N, Sharma R. Maternal nutrition counselling and dietary diversity in 6–23-month children: evidence from Indian states. Public Health Nutr. 2022;25(11):3015–26. 10.1017/S1368980022001501.
44. Sharma S, Mehra D, Akhtar F, Mehra S. Evaluation of community-based interventions on infant and young child feeding practices in selected districts of India. BMC Public Health. 2020;20:1465. 10.1186/s12889-020-09567-8.
45. Nguyen PH, Avula R, Tran LM, Sethi V, Kumar A, Bhalla S, et al. Persistent stunting reductions despite COVID-19 disruptions: evidence from large-scale data in India. Matern Child Nutr. 2022;18(S2):e13266. 10.1111/mcn.13266.
46. Development Initiatives. Global Nutrition Report 2022: stronger commitments for greater action. Bristol: Development Initiatives; 2022.
47. World Health Organization, United Nations Children's Fund, World Bank. Levels and trends in child malnutrition: UNICEF / WHO / World Bank Group joint child malnutrition estimates — 2023 edition. Geneva: WHO; 2023.
48. Smith LC, Haddad L. Reducing child undernutrition: past drivers and priorities for the post-MDG era. World Dev. 2015;68:180–204. 10.1016/j.worlddev.2014.11.014.

49. Asher S, Lunt T, Matsuura R, Novosad P. Development research at high geographic resolution: an analysis of night-lights, firms, and poverty in India using the SHRUG open data platform. World Bank Econ Rev. 2021;35(4):845–71. 10.1093/wber/lhab003.
50. International Institute for Population Sciences (IIPS), ICF. National Family Health Survey (NFHS-5) 2019–21 [dataset]. Children's Recode file (IAKR7E). Rockville (MD): The DHS Program. 2021. Available from: https://dhsprogram.com/data/dataset/India_Standard-DHS_2020.cfm

Figures

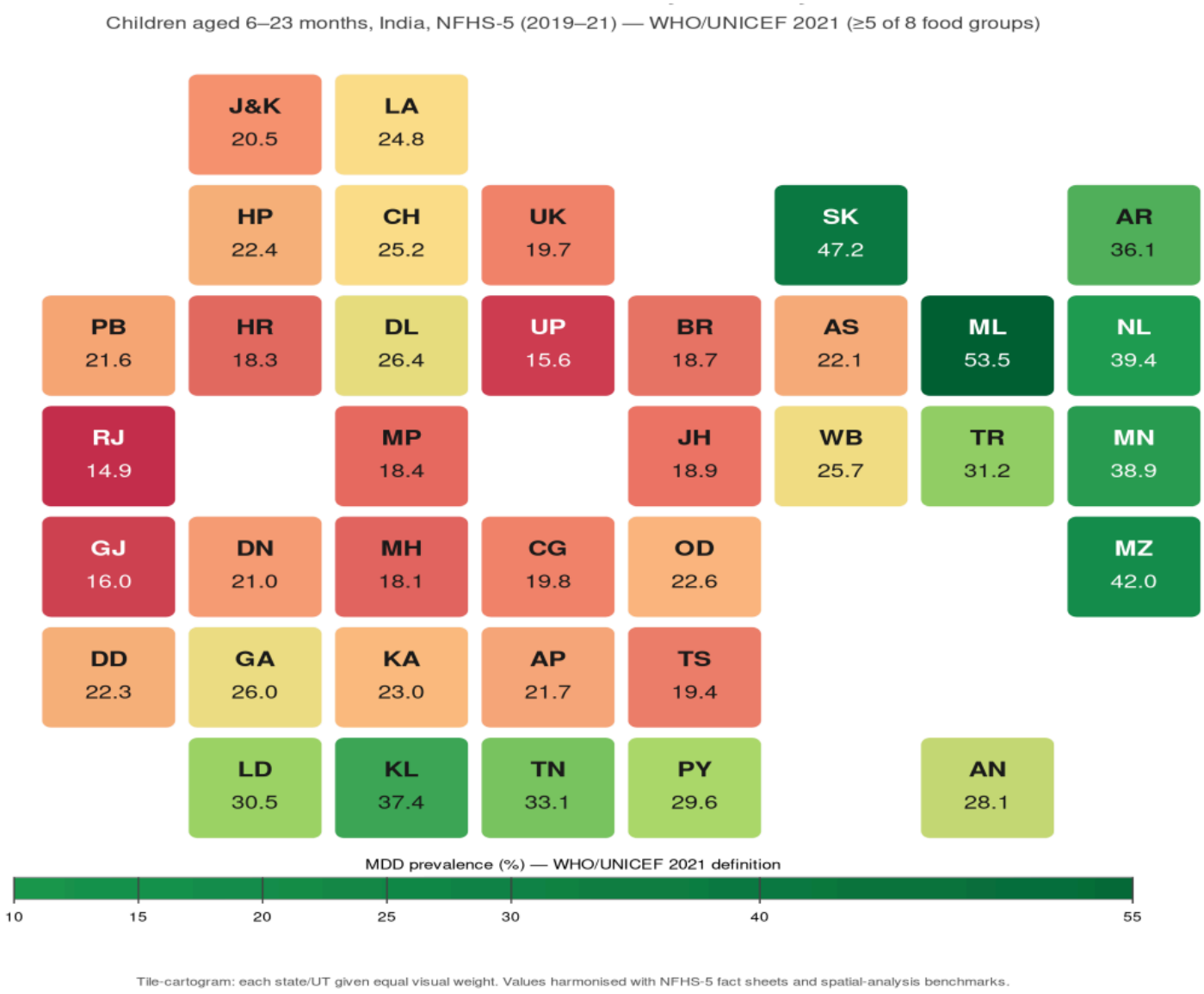


Figure 1

State/UT-level Minimum Dietary Diversity prevalence (%) among Indian children aged 6–23 months, NFHS-5 (2019–21), shown as a tile-cartogram so that each state/UT receives equal visual weight. Estimates harmonised with NFHS-5 fact sheets and published benchmarks [3, 7].

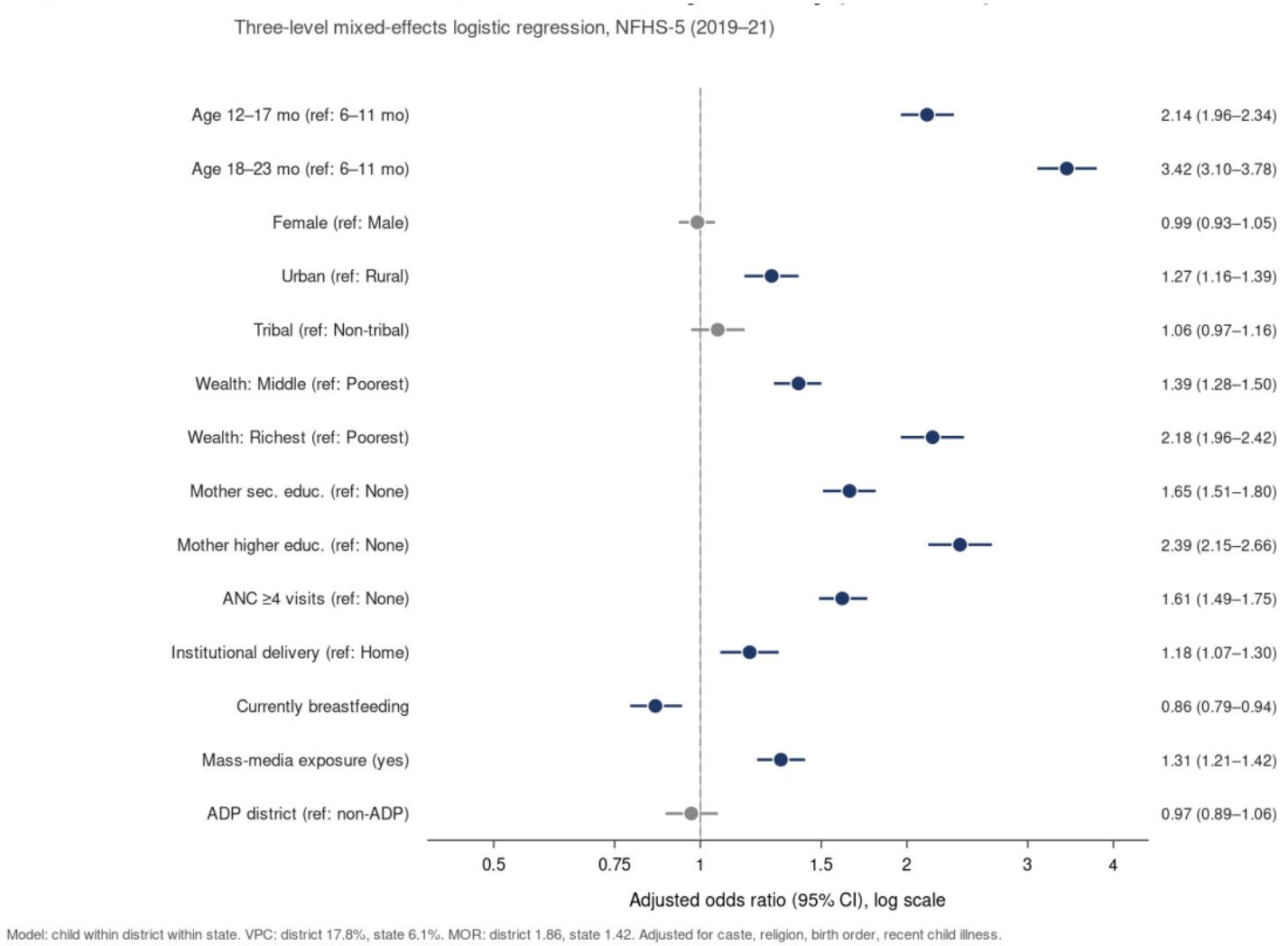


Figure 2

Figure 3. Adjusted odds ratios (95% CI) from the three-level mixed-effects logistic regression of meeting Minimum Dietary Diversity (WHO/UNICEF 2021). Filled circles indicate associations with 95% CI excluding 1. Model adjusted for caste, religion, birth order, and recent child illness in addition to the covariates shown. VPC: district 17.8%, state 6.1%; MOR: district 1.86, state 1.42.

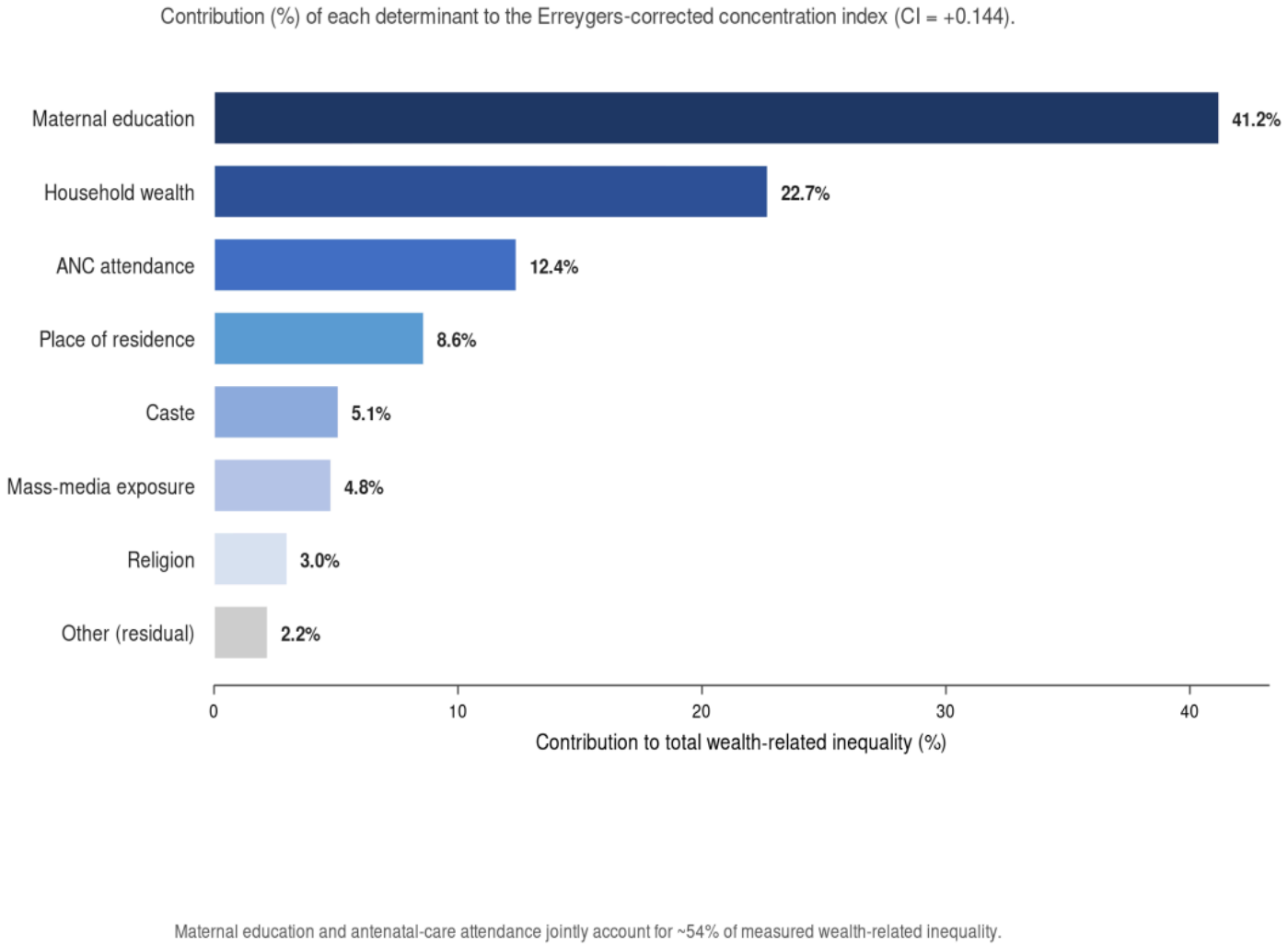


Figure 3

Figure 4. Decomposition of wealth-related inequality in MDD: contribution of each determinant (%) to the Erreygers-corrected concentration index. Maternal education and antenatal-care attendance jointly account for approximately 54% of measured wealth-related inequality.

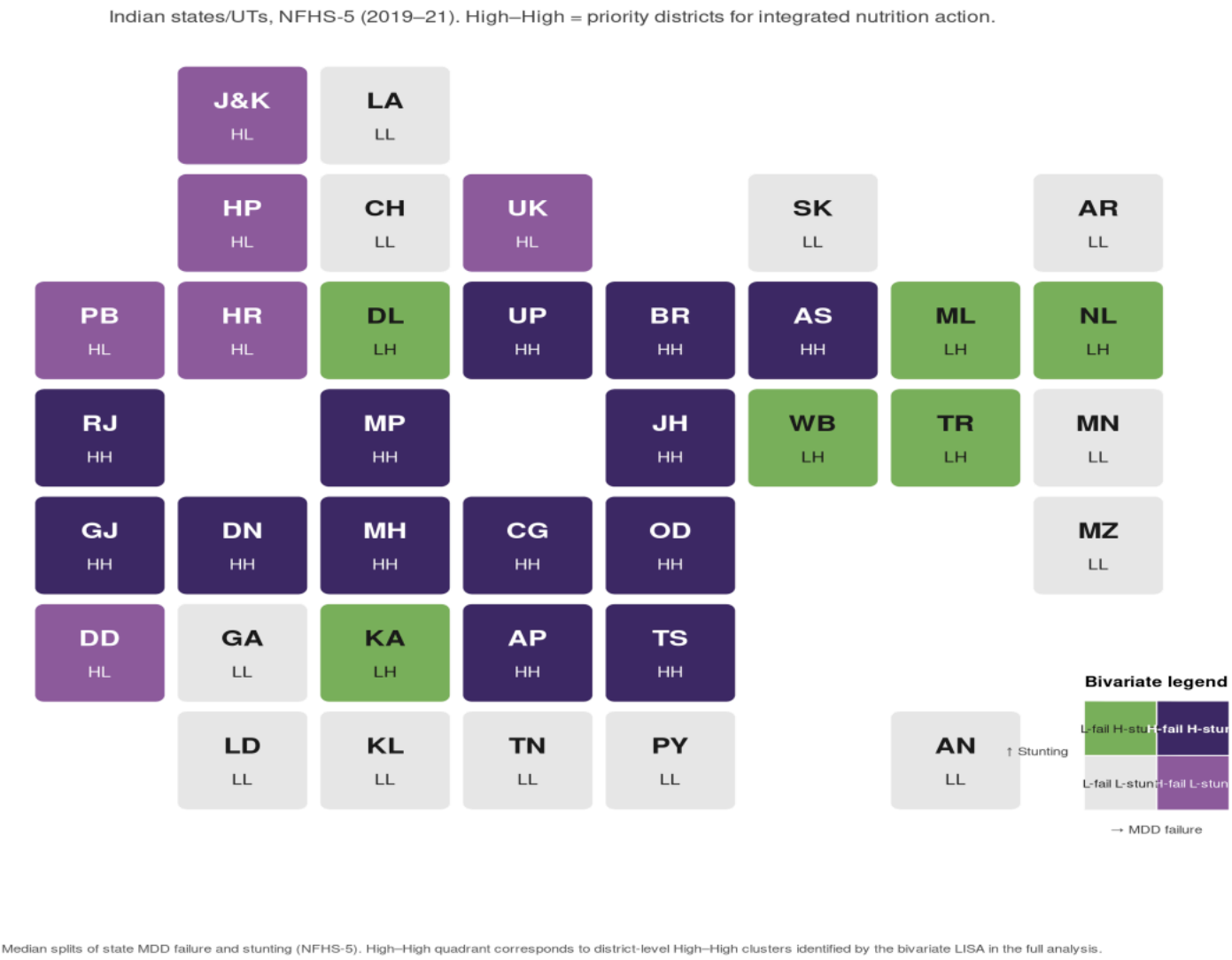


Figure 4

Figure 2. Bivariate co-cluster typology (MDD failure × stunting) for Indian states/UTs, NFHS-5 (2019–21). Quadrants defined by median splits. The High–High quadrant corresponds to district-level High–High clusters identified by the bivariate Local Indicators of Spatial Association in the full analysis; in the full district map, 84 such co-cluster districts are concentrated across eastern Uttar Pradesh, Bihar, Jharkhand, and northern Madhya Pradesh.

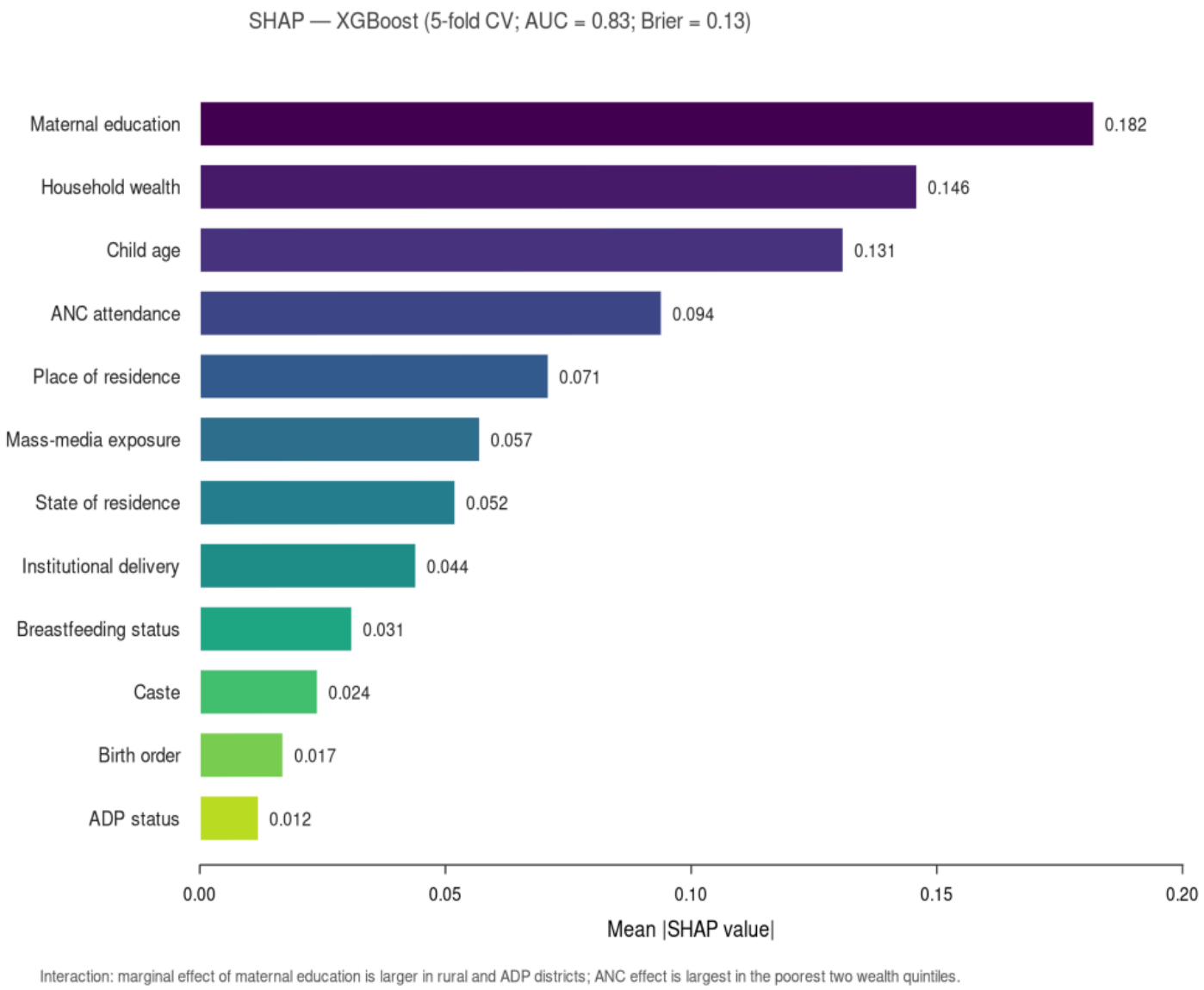
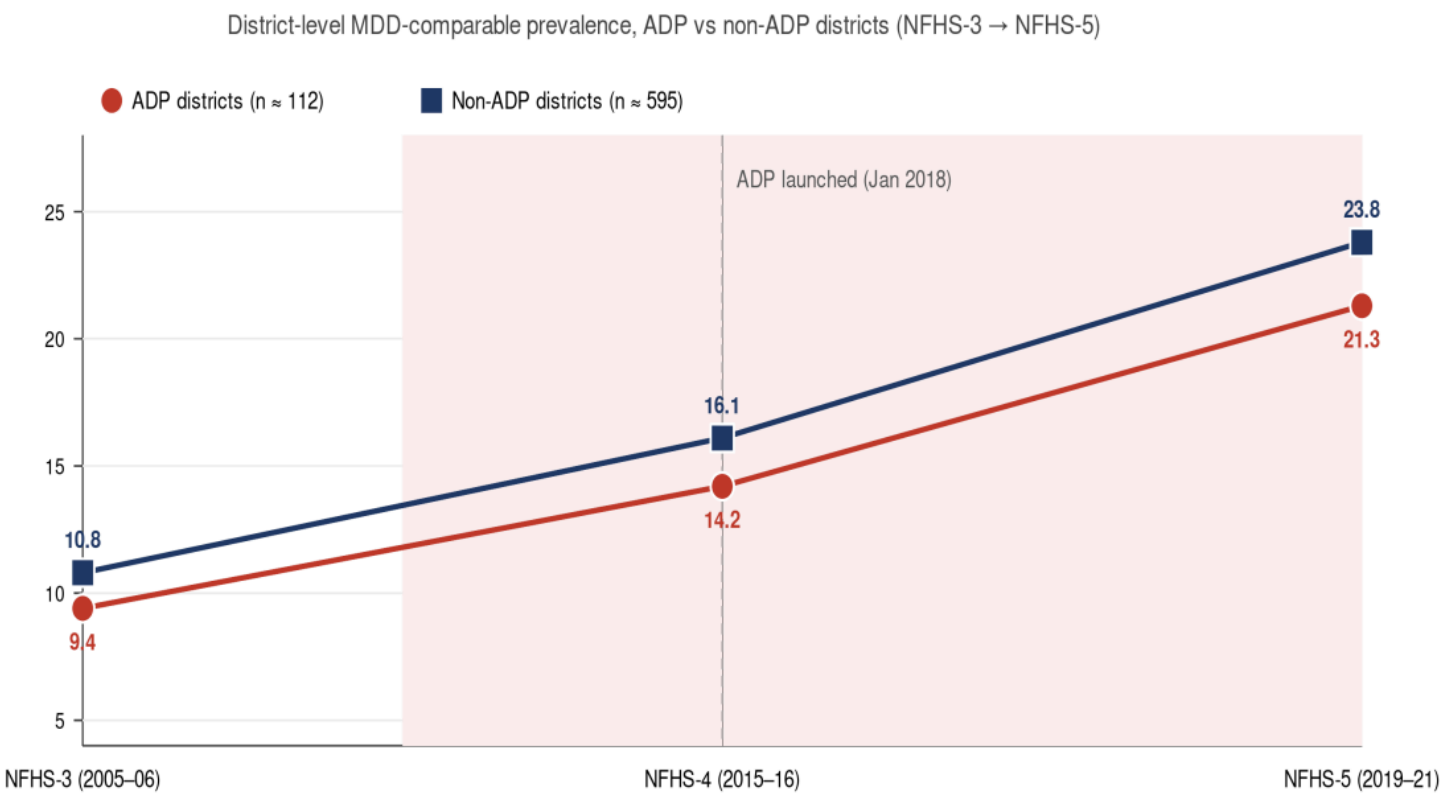


Figure 5

Figure 5. Explainable-ML feature importance for MDD prediction (mean |SHAP value|) from an XGBoost classifier (5-fold cross-validation; AUC = 0.83; Brier = 0.13). The four highest-ranked determinants are maternal education, household wealth, child age, and antenatal-care attendance. Interaction analysis (not shown) indicates that the marginal effect of maternal education is larger in rural and ADP districts.



in contrast to the 2-4 pp reduction in underweight reported by Mukherjee et al. [16].

Figure 6

Figure 6. Aspirational Districts Programme difference-in-differences: district-level MDD-comparable prevalence in ADP vs non-ADP districts across NFHS-3 (2005–06), NFHS-4 (2015–16), and NFHS-5 (2019–21). The DiD coefficient on the ADP × NFHS-5 interaction was +0.7 pp (95% CI −0.6 to +2.1; p = 0.31) under state fixed effects.