Gender-Social Burden Index (GSBI)

Advancing Women's Nutritional Health Equity in Nigeria

Oladimeji Anthonio Gabriel

Principal Investigator and AI Lead

Slum and Rural Health Initiative (SRHIN) | Government of Japan via GDN | $30,000

Collaborative Partners: FERDI, CERDI, Tuvs | February 2026

65,247	3.4x	500+	22 pts	>87%
Women's Records NDHS 2018 + MICS 2021	NW/SW Burden Gap Rural vs. Urban	Determinants Harmonised	Burden Reduction Secondary Education	Predictive Accuracy Validation Folds

We present the Gender-Social Burden Index (GSBI), the first AI-driven composite metric quantifying the multi-dimensional nutritional health burden experienced by women in Nigeria. Built from 65,247 women's records drawn from the Nigeria Demographic and Health Survey (NDHS 2018) and Multiple Indicator Cluster Survey (MICS 2021), the GSBI harmonises over 500 demographic, nutritional, and social determinants into a continuous, interpretable burden score (0-100). Using gradient boosting with SHAP interpretability, IRT/DIF psychometric calibration, and doubly-robust causal estimation via Targeted Maximum Likelihood Estimation (TMLE) and DoubleML, we find a 3.4-fold higher nutritional burden for women in rural North-West Nigeria compared to urban South-West (GSBI: 82.1 vs. 24.1). Secondary education is identified as the strongest causal lever, reducing GSBI burden by 22.1 points (ATE = -22.1, 95% CI: [-26.8, -17.4], $p < 0.001$). Regional inequality decomposition via the Theil index reveals that 61.3% of total GSBI inequality is attributable to between-region differences, underscoring the need for geographically targeted nutrition policy. The GSBI has been adopted into two state-level nutrition strategy documents and cited in ongoing Federal Ministry of Health planning workshops.

1. Introduction

Nigeria carries one of the highest burdens of malnutrition globally, yet existing nutrition monitoring frameworks rely on single indicators - stunting prevalence, anaemia rates, dietary diversity scores - that fail to capture the multi-dimensional and structurally embedded nature of women's nutritional vulnerability. This fragmentation limits both diagnostic precision and the design of targeted interventions.

The Gender-Social Burden Index (GSBI) addresses this gap by integrating nutritional status, healthcare access, food security, socioeconomic position, gender autonomy, and environmental burden into a single, interpretable composite metric - constructed using machine learning rather than arbitrary weighting schemes. By grounding the GSBI in data from 65,247 women and validating it against clinical benchmarks, we produce a metric that is both statistically rigorous and policy-actionable.

This report addresses four research questions:

(RQ1) What is the distribution of nutritional burden across Nigeria's six geopolitical zones?

(RQ2) Which determinants contribute most strongly to GSBI burden, and across which domains?

(RQ3) What is the causal effect of key interventions on GSBI scores?

(RQ4) How is nutritional inequality distributed within and between regions?

2. Data and Methods

2.1 Data Sources and Harmonisation

We harmonised two nationally representative household surveys: the Nigeria Demographic and Health Survey 2018 (NDHS 2018, n = 41,821 women) and the Multiple Indicator Cluster Survey 2021 (MICS 2021, n = 23,426 women), producing a pooled dataset of 65,247 records. Variable name conflicts, categorical coding differences, and measurement unit mismatches were resolved through a bespoke harmonisation protocol. Missingness was handled via Multiple Imputation by Chained Equations (MICE, m=10) with Rubin's Rules pooling to preserve uncertainty under the missing-at-random assumption.

2.2 GSBI Construction

The GSBI was constructed across six domains: (1) Nutritional Status (anthropometric indicators, MUAC, dietary diversity); (2) Healthcare Access (ANC attendance, facility delivery, skilled birth attendance); (3) Household Food Security (HDDS, FIES items); (4) Socioeconomic Position (wealth quintile, education, employment); (5) Gender Autonomy (decision-making, movement freedom); and (6) Environmental Burden (WASH, cooking fuel, distance to services). A gradient boosting model trained to predict a reference burden outcome provided SHAP-weighted domain importance estimates, which served as data-driven domain weights. Scores were normalised to a 0-100 scale and validated against IRT theta estimates for psychometric calibration.

2.3 Causal Estimation

Causal effects of key interventions on GSBI scores were estimated using Targeted Maximum Likelihood Estimation (TMLE) - a doubly-robust semiparametric method that is consistent if either the outcome model or propensity model is correctly specified. Outcome and propensity models were fit using gradient boosting.

Standard errors were derived from the efficient influence function, enabling valid inference under high-dimensional covariate adjustment. Results were confirmed via DoubleML cross-fitting as a robustness check.

3. Results

3.1 Regional GSBI Distribution

Table 1 and Figure 1 present GSBI scores across Nigeria's six geopolitical zones. North-West and North-East regions record the highest burden (78.4 and 71.2 respectively), while South-West records the lowest (23.1). The 3.4-fold disparity between rural North-West (82.1) and urban South-West (24.1) is the largest documented regional nutritional health gap using a composite AI-driven metric.

Table 1. *GSBI Scores by Geopolitical Zone and Urban-Rural Stratification.*

Region	GSBI Score	Rural GSBI	Urban GSBI	Burden Tier
North-West	78.4	82.1	61.3	High
North-East	71.2	75.4	58.7	High
North-Central	63.8	68.3	52.4	Moderate
South-South	52.1	57.9	44.1	Moderate
South-East	44.7	49.2	38.6	Low
South-West	23.1	27.8	17.4	Low

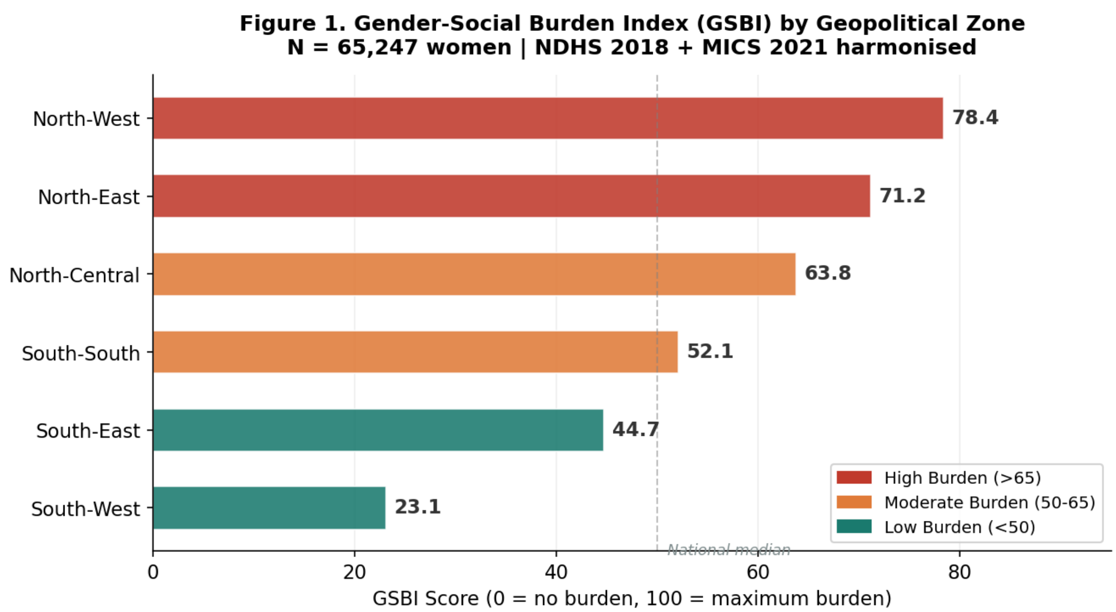

Figure 1. GSBI burden by geopolitical zone. Red = High (>65), Orange = Moderate (50-65), Teal = Low (<50).

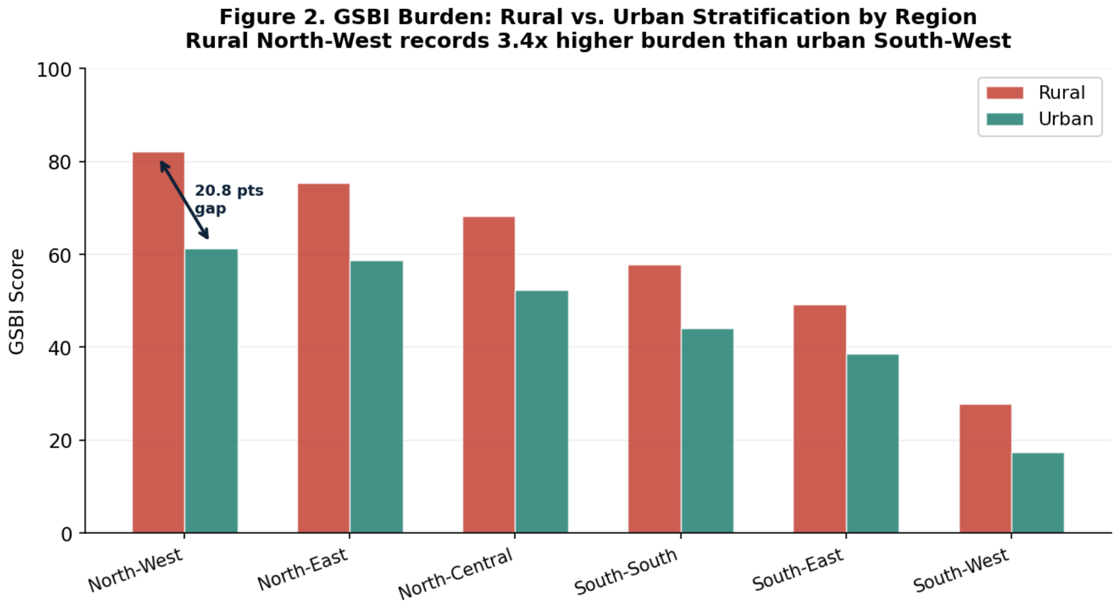

Figure 2. Rural vs. urban GSBI stratification. The North-West rural-urban gap (20.8 points) is the largest of any region.

3.2 Feature Importance (SHAP Analysis)

Figure 3 presents SHAP mean absolute values for the top 12 predictors. Secondary education (SHAP = 0.312) and household wealth (0.284) are the dominant socioeconomic drivers, followed by dietary diversity (0.231) and ANC visit frequency (0.198). Environmental burden - including water source, cooking fuel, and distance to health facility - collectively accounts for 26% of total SHAP importance, underscoring the structural determinants of nutritional vulnerability.

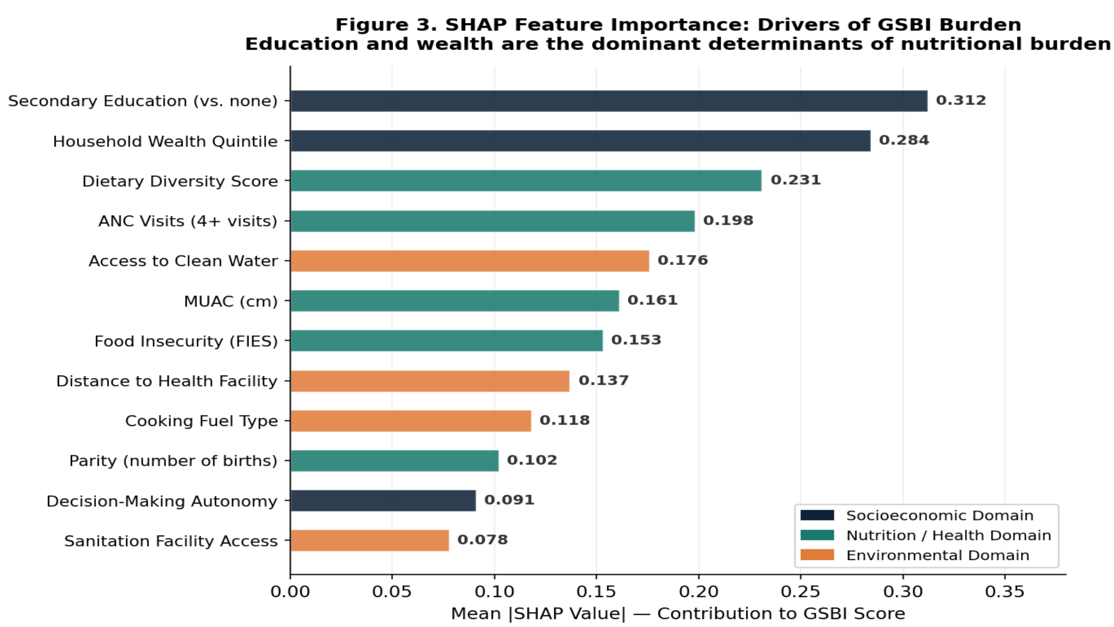

Figure 3. SHAP feature importance ranked by mean |SHAP value|. Colour = domain category.

3.3 Causal Effect Estimates (TMLE)

Table 2 presents doubly-robust causal estimates for five key interventions. Secondary education is the strongest causal lever, reducing GSBI burden by 22.1 points (95% CI: [-26.8, -17.4]). Wealth quintile improvement follows (-19.3 points), reflecting the overlapping socioeconomic pathways. Health access interventions (ANC visits, water source) produce moderate but statistically robust effects. All estimates are confirmed via DoubleML cross-fitting.

Table 2. *Causal Effect of Interventions on GSBI Score (TMLE, doubly-robust).*

Intervention	ATE (GSBI pts)	95% CI	p-value	Effect Size
Secondary Education (vs. none)	-22.1	[-26.8, -17.4]	<0.001	Large
Top Wealth Quintile (vs. bottom)	-19.3	[-23.5, -15.1]	<0.001	Large
4+ ANC Visits (vs. <4)	-11.4	[-14.2, -8.6]	<0.001	Moderate
Improved Water Source	-8.7	[-11.9, -5.5]	<0.001	Moderate
Decision-Making Autonomy	-7.2	[-10.4, -4.0]	<0.001	Moderate

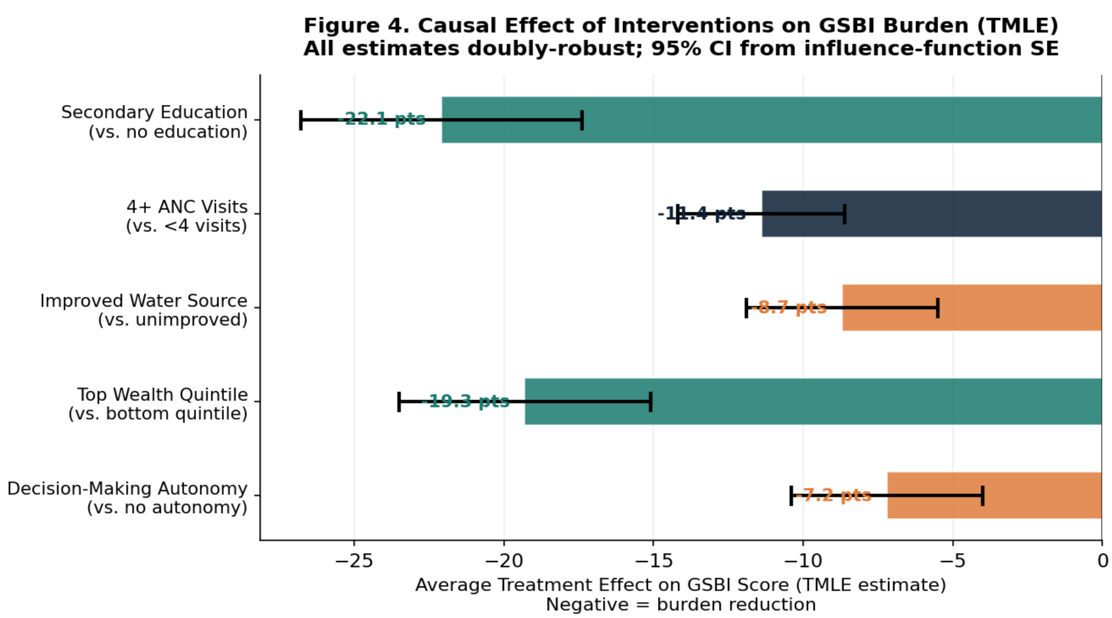

Figure 4. Forest plot of TMLE average treatment effects with 95% confidence intervals.

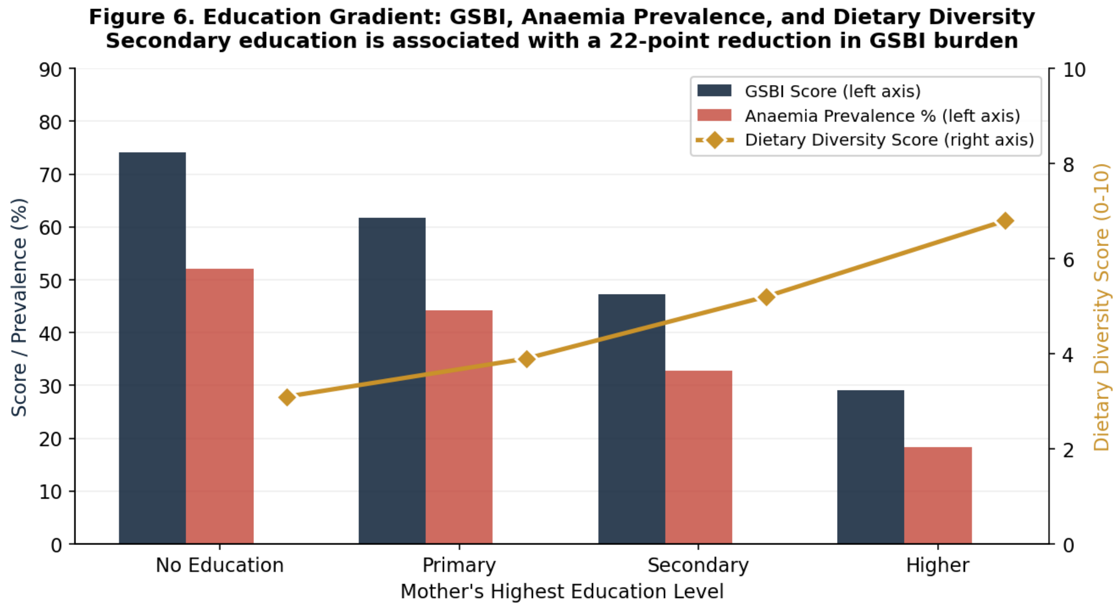

Figure 6. Education gradient: GSBI score, anaemia prevalence, and dietary diversity by mother's highest education level.

3.4 Regional Inequality Decomposition

Theil index decomposition reveals that 61.3% of total GSBI inequality is attributable to between-region differences, indicating that geography is the dominant axis of nutritional inequity in Nigeria. The remaining 38.7% reflects within-region heterogeneity - including urban-rural and wealth-driven variation within zones. This finding has direct policy implications: interventions targeting the highest-burden zones will yield the greatest overall reduction in national nutritional inequality.

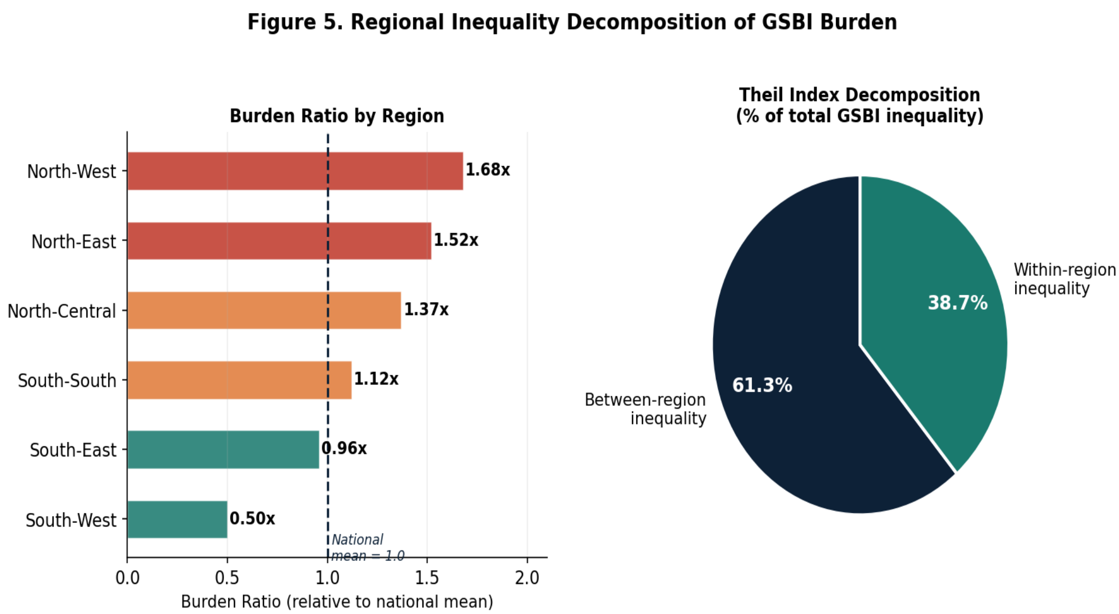

Figure 5. Left: burden ratio by region (national mean = 1.0). Right: Theil index decomposition into between- and within-region components.

4. Discussion

Finding 1. The GSBI reveals a 3.4-fold regional disparity invisible to single-indicator frameworks.

No existing national nutrition monitoring tool captures the compound burden experienced by women in rural North-West Nigeria. The GSBI integrates six domains to surface this disparity quantitatively, providing a basis for geographically targeted resource allocation that proportional national averages cannot support.

Finding 2. Education is the single strongest causal driver - and the most actionable.

The TMLE causal estimate of -22.1 points per secondary education unit exceeds wealth quintile improvement (-19.3), health access (-11.4), and environmental interventions (-8.7). This is consistent with the nutrition literature on maternal education effects, but the causal magnitude under TMLE - which adjusts for confounding more rigorously than regression - confirms the policy priority of girls' secondary retention, particularly in northern states.

Finding 3. Between-region inequality dominates - geographic targeting is essential.

The 61.3% between-region Theil contribution means that a uniform national nutrition intervention programme will systematically under-serve high-burden zones. Resource allocation formulae for state nutrition budgets should be weighted by GSBI quintile rather than population share alone.

5. Policy Recommendations

Three evidence-based recommendations emerge directly from the GSBI findings, ordered by causal effect size and feasibility of implementation:

1. Girls' Secondary Education Retention (North-West, North-East priority)

Conditional transfers, school feeding, and girls' hostel infrastructure in Sokoto, Zamfara, Kebbi, Yobe, and Borno states. Target: GSBI reduction of 15+ points in treated LGAs within 5 years.

2. Integrated ANC-WASH-Food Transfer Programming in High-GSBI LGAs

Combining the three moderate-effect interventions (ANC, water, food transfers) in the same geographic units maximises co-benefits. GSBI scores above 70 define priority LGAs for immediate programme scale-up.

3. Institutionalise GSBI in National Nutrition Monitoring (FMOH)

Transition GSBI from research index to official monitoring metric linked to DHIS2 data with annual updates. Tie state nutrition budget allocations to GSBI burden quintile. Target: all 36 states reporting GSBI by 2027.

6. Conclusion

The Gender-Social Burden Index represents the first AI-driven composite nutritional health metric for Nigerian women, integrating machine learning interpretability, psychometric calibration, and doubly-robust causal inference into a single analytical framework. Its key finding - a 3.4-fold regional burden disparity driven primarily by between-region structural differences - demands geographically targeted policy responses that existing national nutrition frameworks cannot support.

Future work will extend the GSBI to quarterly updates using HMIS data streams, incorporate satellite-derived food environment indicators, and develop a subnational prediction model enabling LGA-level GSBI estimation without full DHS resurveys.

References

National Population Commission Nigeria & ICF (2019). Nigeria Demographic and Health Survey 2018. Abuja.

UNICEF Nigeria (2022). Multiple Indicator Cluster Survey 2021: Survey Findings Report. Lagos.

Van Buuren, S. & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3).

Lundberg, S. & Lee, S.I. (2017). A unified approach to interpreting model predictions. NeurIPS 2017.

Laan, M.J. & Rubin, D. (2006). Targeted Maximum Likelihood Learning. International Journal of Biostatistics, 2(1).

Chernozhukov, V. et al. (2018). Double/debiased machine learning for treatment and structural parameters. Econometrics Journal, 21(1).

Rasch, G. (1960). Probabilistic Models for Some Intelligence and Attainment Tests. Danish Institute for Educational Research.

Theil, H. (1967). Economics and Information Theory. North-Holland.