LINK MEDICAL JOURNAL
of Health and Community Research | LMJHCR
Volume 4, Issue 1 | June 2026 | ISSN 3007-3448

Indexed in DOAJ | CC BY 4.0
DOI: 10.61919/thq1qy88

Original Article

Geographic and Socioeconomic Predictors of Postpartum Depression Identified Through Machine Learning Analysis of Pakistani Health Data

Sabira Feroz[1], Sidra Irfan[2], Fariha Muhammad Iqbal[3], Noorulain Tariq Siddiqui[4], Zubia Shahid[5], Saeed Wali[6], Iraj Siddiqui[7]

[1] Senior Lecturer, Bahria University, Islamabad, Pakistan
[2] Senior Registrar, Preventive Medicine, Amiri Hospital, Kuwait City, Kuwait
[3] Doctor, Muhammad College of Medicine, Peshawar, Pakistan
[4] BSc (Hons), Prosthetic and Orthotic Sciences, PIPOS, Khyber Medical University, Peshawar, Pakistan; Quality Assurance Manager, Vital Medical Associates, PC, USA
[5] Lecturer, Quaid-i-Azam University, Islamabad, Pakistan
[6] MBBS, Bolan Medical College, Quetta, Pakistan
[7] MBBS, ZMC, Karachi, Pakistan; House Officer, Ziauddin University Hospital, Karachi, Pakistan

***Corresponding author: Sabira Feroz, sabira.buic@bahria.edu.pk**

"Cite this Article" | **Received**: 11 January 2026; **Accepted**: 30 May 2026; **Published**: 30 June 2026.

ABSTRACT

Background: Postpartum depressive symptoms are frequently under-recognised in resource-constrained settings, and their geographic and socioeconomic distribution within Pakistani hospital populations remains insufficiently characterised. Objective: To identify hospital-city and socioeconomic predictors of postpartum depressive symptom severity using conventional regression and random forest analysis. Methods: This multicentre retrospective record-based analytical study included 336 postpartum women aged 18–45 years receiving care at tertiary hospitals in Karachi, Lahore, and Rawalpindi between January 2021 and December 2023. The primary outcome was the continuous Edinburgh Postnatal Depression Scale score. Geographic and socioeconomic predictors were examined using analysis of variance, multivariable linear regression, and random forest regression. Results: The mean EPDS score was 11.2 ± 5.8, and 187 women (55.7%) screened positive at an EPDS threshold of ≥10. Mean scores were highest in Rawalpindi (13.4 ± 5.9), followed by Lahore (11.1 ± 5.6) and Karachi (9.1 ± 5.2). Compared with the respective reference categories, Rawalpindi hospital-city site was associated with a 3.9-point higher score, household income below PKR 30,000 with a 5.8-point increase, no formal education with a 4.9-point increase, and overcrowding with a 3.2-point increase. The random forest model achieved a test-set R^2 of 0.51. Conclusion: Socioeconomic disadvantage and hospital-city context were associated with greater postpartum depressive symptom severity. Targeted screening may be warranted among women from low-income, poorly educated, and overcrowded households. Keywords: Edinburgh Postnatal Depression Scale; geographic variation; household income; machine learning; maternal education; overcrowding; Pakistan; postpartum depression.

EDITORIAL INFORMATION

Author Contributions: Concept: SF, SI; Design: SF, FMI, NTS; Data Collection: SI, FMI, ZS, SW, IS; Analysis: SF, NTS, ZS; Drafting: SF, SI, FMI, NTS, ZS, SW, IS.
Ethical Approval: Bahria University, Islamabad, Pakistan.
Informed Consent: Written informed consent was obtained from all participants
Conflict of Interest: The authors declare no conflict of interest; **Funding**: No external funding; **Data Availability**: Available from the corresponding author on reasonable request; **Acknowledgments**: N/A.

INTRODUCTION

Postpartum depression is a common but frequently under-recognised complication of childbirth that can adversely affect maternal functioning, infant development, family relationships, and engagement with postnatal healthcare. The burden is particularly important in low- and middle-income countries, where limited mental-health resources, inadequate routine screening, social stigma, and fragmented referral pathways may delay recognition and treatment. In Pakistan, postpartum depressive symptoms may be further influenced by socioeconomic disadvantage, restricted access to healthcare, limited mental-health literacy, and social conditions that affect women during pregnancy and the early postpartum period (1).

Previous research from Pakistan and other South Asian settings has identified several individual and household factors associated with postpartum depressive symptoms, including low household income, limited maternal education, intimate partner violence, inadequate social support, and adverse living conditions. However, much of the available evidence has been derived from single-centre or single-city studies, limiting assessment of whether

the observed burden differs across hospital catchment areas and socioeconomic contexts (2,3). Geographic comparisons may be clinically useful because women receiving care in different cities and hospitals can vary in referral patterns, healthcare access, socioeconomic circumstances, cultural practices, and availability of postnatal mental-health services. Nevertheless, hospital-based geographic differences must be interpreted cautiously because the effects of city, hospital characteristics, and catchment population may overlap.

Conventional regression methods remain important for estimating adjusted associations and expressing the direction and magnitude of relationships between socioeconomic factors and postpartum depressive symptom severity. These methods may, however, have limited ability to identify complex non-linear patterns and higher-order relationships unless these are specified in advance. Machine-learning approaches such as random forest regression provide a complementary predictive framework by allowing multiple demographic, geographic, socioeconomic, and obstetric characteristics to be considered simultaneously without requiring all relationships to be modelled as linear (4–6). Such techniques can rank predictors according to their contribution to prediction, although variable importance should not be interpreted as evidence of causality.

A more precise understanding of the characteristics associated with postpartum depressive symptom severity may help clinicians identify women who require closer assessment during routine postnatal care. This is relevant in resource-constrained settings, where universal access to specialist psychiatric assessment may not be feasible and pragmatic screening strategies are needed. Evidence derived from routinely collected clinical data may also support the development of locally relevant risk-stratification approaches, provided that predictive findings are interpreted within the limitations of hospital-based sampling and are subsequently validated in independent populations (7–9).

The present study therefore examined geographic and socioeconomic predictors of postpartum depressive symptom severity among women receiving postnatal care at three tertiary hospitals in Karachi, Lahore, and Rawalpindi. Specifically, it evaluated differences in Edinburgh Postnatal Depression Scale scores across the three hospital-city sites and assessed the associations of household income, maternal education, employment status, housing conditions, and household overcrowding with symptom severity. Multivariable linear regression was used to estimate adjusted associations, whereas random forest regression was applied as a complementary predictive method to rank the relative contribution of the included characteristics. It was hypothesised that socioeconomic disadvantage, particularly low income, limited education, and household overcrowding, would be associated with higher postpartum depressive symptom severity after adjustment for relevant maternal and obstetric characteristics.

MATERIALS AND METHODS

This multicentre retrospective record-based analytical study used pre-existing clinical and administrative data to identify geographic and socioeconomic characteristics associated with postpartum depressive symptom severity. Records were obtained from three tertiary care hospitals located in Karachi, Lahore, and Rawalpindi. These hospitals served urban and surrounding catchment populations and were selected because they had high obstetric patient volumes, routinely documented socioeconomic information, and used a standardised postpartum mental-health screening measure. Because one participating hospital represented each city, the hospital site and city-level category were not analytically separable; therefore, geographic comparisons represented differences between the three hospital-city sites rather than independent population-level city effects. Medical records from January 1, 2021, to December 31, 2023, were reviewed. Data extraction, cleaning, harmonisation, and analysis were undertaken from January to September 2024. No prospective intervention or follow-up was performed.

The target population comprised postpartum women aged 18–45 years who had delivered a live singleton infant and received postnatal care at one of the participating hospitals. Women were eligible when a completed Edinburgh Postnatal Depression Scale assessment was documented between 24 hours and six weeks after delivery, their place of residence could be assigned to the relevant hospital catchment area, and information was available on the principal socioeconomic variables, including household income, maternal education, and employment status. Women were excluded when their records documented major depressive disorder, bipolar disorder, psychotic disorder, or treatment for a serious pre-existing psychiatric condition before the index pregnancy; when the pregnancy resulted in stillbirth or neonatal death during the index admission; when more than

two EPDS items were missing; or when the available hospital documentation did not permit use of the record for research purposes. The primary analysis focused on postpartum depressive symptom severity rather than a confirmed psychiatric diagnosis.

The initial sample-size target was 350 records, allowing for incomplete information and exclusions, and the final analytical sample comprised 336 women. Within each hospital, potentially eligible records were arranged chronologically according to delivery date. Systematic sampling was then applied separately at each site, and every third eligible record was selected until the hospital-specific allocation, determined in proportion to average delivery volume, was reached. When a selected record was unavailable for review, the next chronologically available eligible record was assessed and the systematic sequence was subsequently continued. The sampling procedure was applied independently at each hospital to maintain representation across the three sites.

Data were extracted from electronic medical records and paper-based obstetric, postnatal, and psychiatric screening records using a standardised case-report form. Two trained research assistants independently reviewed the source records. Extracted entries were compared, and discrepancies were resolved through re-examination of the original documentation. Personal identifiers, medical record numbers, and detailed addresses were removed from the analytical dataset and replaced with unique study codes before data cleaning and analysis.

The primary outcome was postpartum depressive symptom severity, measured using the total Edinburgh Postnatal Depression Scale score. The EPDS is a 10-item self-report screening instrument with a total score ranging from 0 to 30, with higher scores indicating greater depressive symptom severity (10). The continuous total score was used for the primary analyses because the study aimed to examine variation in symptom severity. For secondary descriptive analyses, an EPDS score of 10 or higher was classified as probable postpartum depression. This classification represented a screening threshold and was not considered equivalent to a clinical diagnosis. Only assessments documented as having been administered by trained nursing or midwifery personnel according to the hospitals' routine procedures were included. Records without a documented assessment time or identifiable administering healthcare professional were excluded. The interval between delivery and EPDS administration was retained during data extraction because assessments could occur between 24 hours and six weeks postpartum.

The principal explanatory variable was hospital-city site, categorised as Karachi, Lahore, or Rawalpindi. Because each city was represented by a single hospital, this variable incorporated both geographic context and potential hospital-level differences. Socioeconomic variables included monthly household income, categorised as less than PKR 30,000, PKR 30,000–60,000, more than PKR 60,000–100,000, and more than PKR 100,000; maternal and paternal education, classified as no formal education, primary, secondary, or higher secondary and above; maternal employment status, classified as housewife, employed outside the home, or self-employed; and housing status, categorised as owned, rented, or informal settlement/katchi abadi. Household overcrowding was defined as more than two persons per habitable room, excluding kitchens and bathrooms.

Maternal and obstetric characteristics selected a priori as potential confounders included maternal age, parity, mode of delivery, previous pregnancy loss, neonatal admission to a special-care unit, and documented intimate partner violence during the index pregnancy. Mode of delivery was classified as vaginal, caesarean, or assisted delivery. Information on intimate partner violence was extracted from routine antenatal records when available. Because this variable had a higher proportion of missing observations than the other covariates, it was excluded from the primary adjusted model and examined separately in sensitivity analyses.

Several procedures were used to reduce bias and improve data quality. All potentially eligible records were identified before sampling at each site. The same eligibility criteria, sampling sequence, extraction form, coding framework, and variable definitions were applied across the three hospitals. EPDS results were accepted only when the record documented administration by an appropriate healthcare professional and provided sufficient information to establish the timing of assessment. Independent extraction and verification by two trained reviewers were used to minimise transcription and classification errors. Nevertheless, because the study relied on routinely collected hospital records, differences in documentation practices and referral characteristics across sites could not be eliminated.

The completeness and distribution of each variable were examined before analysis. Records missing more than two EPDS items were excluded because a reliable total score could not be determined. For maternal age, parity, and mode of delivery, which had less than 5% missing data, analyses were based on available complete observations. Missing values for household income and education were addressed through multiple imputation by chained equations, using five imputed datasets under the assumption that missingness was conditionally related to observed hospital, housing, demographic, and obstetric characteristics. The imputation framework incorporated the principal outcome, hospital-city site, socioeconomic variables, and relevant covariates. Estimates from the linear-regression analyses were combined across imputed datasets using standard pooling procedures. Intimate partner violence, for which approximately 12% of values were missing, was excluded from the primary model and evaluated in sensitivity analyses using records with available information.

Statistical analyses were performed using R version 4.3.2. Continuous variables were summarised as means and standard deviations when their distributions were approximately symmetric, whereas categorical variables were presented as frequencies and percentages. The distribution of EPDS scores was assessed through graphical examination and evaluation of model residuals. Differences in participant characteristics across hospital-city sites were assessed using the chi-square test for categorical variables. Mean EPDS scores across the three hospital-city categories were compared using one-way analysis of variance, followed by Tukey's honestly significant difference procedure for pairwise comparisons. Associations between EPDS scores and individual socioeconomic variables were initially examined using independent-samples t tests for binary variables and one-way analysis of variance for variables with more than two categories.

A multivariable linear-regression model was subsequently fitted to estimate adjusted associations between hospital-city site, socioeconomic characteristics, and EPDS symptom severity. The model included household income, maternal education, maternal employment, housing status, household overcrowding, maternal age, parity, mode of delivery, history of pregnancy loss, and neonatal admission to special care. Karachi, the highest household-income category, higher secondary education or above, housewife status, owned housing, and absence of overcrowding were used as reference categories where applicable. Regression coefficients were expressed as adjusted mean differences in EPDS points with 95% confidence intervals and two-sided p-values. Model assumptions were assessed using residual plots, evaluation of residual distribution, examination of influential observations, and assessment of collinearity among predictors. Statistical significance was set at $p \leq 0.05$.

Random forest regression was used as a complementary predictive analysis rather than as a method for causal inference. The available data were partitioned into training and held-out test sets in an 80:20 ratio, with stratification by hospital-city site to preserve representation from all three locations. All geographic, socioeconomic, maternal, and obstetric variables included in the primary analytical framework were considered candidate predictors. Hyperparameters were selected through ten-fold cross-validation within the training data. Predictive performance was evaluated in the held-out test set using root mean square error and R^2. Variable-importance estimates were used to rank the contribution of the included predictors to EPDS score prediction. These rankings were interpreted as measures of predictive contribution within the analysed dataset and not as estimates of independent causal effects. The random forest results were considered alongside the multivariable regression estimates to determine whether the principal predictive patterns were consistent across conventional and machine-learning approaches.

All extracted data were de-identified before analysis, and access to the analytical dataset was restricted to the research team. Names, medical record numbers, and detailed residential identifiers were removed and replaced with study codes. No direct contact with participants occurred, and no intervention was administered. Data were used only for the stated research purpose and were handled according to established principles of confidentiality, data minimisation, and secure processing of sensitive health information.

RESULTS

Following record screening, data cleaning, and application of the eligibility criteria, 336 postpartum women were included in the analysis. Their mean age was 28.4 ± 5.1 years, with an observed range of 18–43 years. Representation across the three hospital-city sites was balanced: 112 women (33.3%) were recruited from

Karachi, 114 (33.9%) from Lahore, and 110 (32.7%) from Rawalpindi. Participant socioeconomic characteristics are summarised in Table 1.

Table 1. Demographic and Socioeconomic Characteristics of the Participants

Characteristic	Category	n (%)
Total sample		336 (100.0)
Age, years	Mean ± SD	28.4 ± 5.1
Hospital-city site	Karachi	112 (33.3)
	Lahore	114 (33.9)
	Rawalpindi	110 (32.7)
Maternal education	No formal education	57 (17.0)
	Primary education	78 (23.2)
	Secondary education or above	201 (59.8)
Monthly household income, PKR	<30,000	94 (28.0)
	30,000–60,000	126 (37.5)
	>60,000–100,000	78 (23.2)
	>100,000	38 (11.3)
Housing type	Owned	184 (54.8)
	Rented	104 (31.0)
	Informal settlement/katchi abadi	48 (14.3)
Household overcrowding	Present	142 (42.3)
	Absent	194 (57.7)

Values are presented as n (%) unless otherwise indicated.

Most participants had completed at least secondary education (n = 201, 59.8%), whereas 57 women (17.0%) had no formal education. The largest household-income category was PKR 30,000–60,000 per month, comprising 126 participants (37.5%), followed by income below PKR 30,000 among 94 women (28.0%). Most women lived in owned housing (n = 184, 54.8%), while 48 (14.3%) lived in informal settlements. Household overcrowding was documented in 142 households (42.3%).

Socioeconomic characteristics differed significantly among the three hospital-city sites. Household income distribution ($\chi^2 = 42.1$, $p < 0.001$), maternal education ($\chi^2 = 38.7$, $p < 0.001$), and housing type ($\chi^2 = 29.4$, $p < 0.001$) varied across Karachi, Lahore, and Rawalpindi. Selected site-specific findings are presented in Table 2.

Table 2. Selected Socioeconomic Characteristics Across Hospital-City Sites

Characteristic	Karachi, n/N (%)	Lahore, n/N (%)	Rawalpindi, n/N (%)	Test statistic	p-value
Monthly income <PKR 30,000	—	—	41/110 (37.3)	$\chi^2 = 42.1$	<0.001
Monthly income >PKR 100,000	21/112 (18.8)	—	—	—	—
Employed outside the home	28/112 (25.0)	24/114 (21.1)	11/110 (10.0)	—	—
Intimate partner violence*	—	—	24/97 (24.7)	—	—

Intimate partner violence data were available for 296 women. Dashes indicate that the corresponding site-specific values or test results were not reported in the source manuscript. The reported omnibus chi-square test for income compared the full income-category distribution rather than only the categories displayed in this abbreviated table.

Rawalpindi had the largest reported proportion of households with monthly income below PKR 30,000 (41/110, 37.3%), whereas Karachi had the largest reported proportion with income above PKR 100,000 (21/112, 18.8%). Employment outside the home was least frequent in Rawalpindi, where 11 of 110 women (10.0%) were employed, compared with 28 of 112 women in Karachi (25.0%) and 24 of 114 women in Lahore (21.1%). Information on intimate partner violence was available for 296 participants, of whom 58 (19.6%) had documented exposure during the index pregnancy. Rawalpindi had the highest reported site-specific frequency, affecting 24 of 97 women with available data (24.7%).

The mean EPDS score in the complete sample was 11.2 ± 5.8. Mean scores differed across hospital-city sites and socioeconomic categories, as shown in Table 3. Rawalpindi had the highest mean EPDS score, followed by Lahore and Karachi. A clear socioeconomic gradient was also observed across household-income and maternal-education categories. A progressive inverse relationship was observed between household income and depressive symptom severity. Women in households earning less than PKR 30,000 had a mean EPDS score of 14.2 ± 5.5, compared with 6.9 ± 4.6 among women in households earning more than PKR 100,000. The unadjusted difference between these two income categories was 7.3 EPDS points. Women with no formal education had a mean score

of 14.8 ± 5.7, compared with 9.5 ± 5.3 among those with higher secondary education or above, representing an unadjusted difference of 5.3 points. The mean EPDS score was also 4.5 points higher among women living in overcrowded households than among those in non-overcrowded households.

Table 3. Edinburgh Postnatal Depression Scale Scores by Geographic and Socioeconomic Characteristics

Characteristic	Category	n	EPDS score, mean ± SD
Complete sample	—	336	11.2 ± 5.8
Hospital-city site	Karachi	112	9.1 ± 5.2
	Lahore	114	11.1 ± 5.6
	Rawalpindi	110	13.4 ± 5.9
Monthly household income, PKR	<30,000	94	14.2 ± 5.5
	30,000–60,000	126	11.8 ± 5.3
	>60,000–100,000	78	9.4 ± 5.1
	>100,000	38	6.9 ± 4.6
Maternal education	No formal education	57	14.8 ± 5.7
	Higher secondary education or above	—	9.5 ± 5.3
Household overcrowding	Present	142	13.9 ± 5.6
	Absent	194	9.4 ± 5.2

EPDS: Edinburgh Postnatal Depression Scale; PKR: Pakistani rupees; SD: standard deviation. The number of women with higher secondary education or above was not reported separately in the source manuscript.

Using an EPDS threshold of 10 or higher, 187 of 336 women (55.7%) screened positive for probable postpartum depression. The corresponding proportions were 43.8% in Karachi, 56.1% in Lahore, and 67.3% in Rawalpindi. Based on the reported site denominators and percentages, these proportions corresponded to approximately 49 women in Karachi, 64 in Lahore, and 74 in Rawalpindi, which together accounted for the reported 187 positive screens.

Table 4. Probable Postpartum Depression by Hospital-City Site

Hospital-city site	Total n	EPDS ≥10, n (%)
Karachi	112	49 (43.8)
Lahore	114	64 (56.1)
Rawalpindi	110	74 (67.3)
Overall	336	187 (55.7)

Counts for individual sites were derived from the reported percentages and denominators and reconcile exactly with the reported overall total. EPDS: Edinburgh Postnatal Depression Scale.

Mean EPDS scores differed significantly among the three hospital-city sites (one-way ANOVA: $F[2,333] = 16.8$, $p < 0.001$). Tukey-adjusted pairwise comparisons showed that Rawalpindi had a mean score 4.3 points higher than Karachi (95% CI: 2.6 to 6.0, $p < 0.001$) and 2.3 points higher than Lahore (95% CI: 0.6 to 4.0, $p = 0.005$). The 2.0-point difference between Lahore and Karachi was not statistically significant (95% CI: −0.3 to 4.3, $p = 0.090$).

Table 5. Pairwise Comparisons of Mean EPDS Scores Across Hospital-City Sites

Comparison	Mean difference, EPDS points	95% CI	p-value
Rawalpindi vs Karachi	4.3	2.6 to 6.0	<0.001
Rawalpindi vs Lahore	2.3	0.6 to 4.0	0.005
Lahore vs Karachi	2.0	−0.3 to 4.3	0.090

Pairwise comparisons were performed using Tukey's honestly significant difference procedure. A positive estimate indicates a higher mean score in the first-listed hospital-city site. CI: confidence interval; EPDS: Edinburgh Postnatal Depression Scale.

Categorical analyses using the EPDS threshold of 10 or higher demonstrated significant associations between probable postpartum depression and household income ($\chi^2 = 31.4$, $df = 3$, $p < 0.001$), maternal education ($\chi^2 = 28.9$, $df = 3$, $p < 0.001$), household overcrowding ($\chi^2 = 18.2$, $df = 1$, $p < 0.001$), and housing type ($\chi^2 = 12.7$, $df = 2$, $p = 0.002$). Maternal employment status was not significantly associated with probable postpartum depression ($\chi^2 = 2.1$, $df = 2$, $p = 0.350$).

In the multivariable linear-regression analysis, hospital-city site and socioeconomic disadvantage remained independently associated with EPDS symptom severity. Compared with Karachi, residence within the Rawalpindi

hospital catchment was associated with an adjusted 3.9-point higher EPDS score (95% CI: 2.3 to 5.5, p < 0.001). Women in the lowest household-income category had an adjusted score 5.8 points higher than those in the highest-income category (95% CI: 3.9 to 7.7, p < 0.001). No formal maternal education was associated with a 4.9-point increase compared with the highest education category (95% CI: 3.1 to 6.7, p < 0.001), while household overcrowding was associated with a 3.2-point increase compared with non-overcrowded housing (95% CI: 1.8 to 4.6, p < 0.001).

Table 6. Associations Between Socioeconomic Characteristics and Probable Postpartum Depression

Characteristic	χ^2	df	p-value
Monthly household income	31.4	3	<0.001
Maternal education	28.9	3	<0.001
Household overcrowding	18.2	1	<0.001
Housing type	12.7	2	0.002
Maternal employment status	2.1	2	0.350

Probable postpartum depression was defined as an EPDS score ≥10. EPDS: Edinburgh Postnatal Depression Scale.

Table 7. Principal Adjusted Predictors of EPDS Symptom Severity

Predictor	Comparison/ reference category	Mean difference, EPDS points	95% CI	p-value
Hospital-city site	Rawalpindi vs Karachi	3.9	2.3 to 5.5	<0.001
Monthly household income	<PKR 30,000 vs >PKR 100,000	5.8	3.9 to 7.7	<0.001
Maternal education	No formal education vs highest education category	4.9	3.1 to 6.7	<0.001
Household overcrowding	Present vs absent	3.2	1.8 to 4.6	<0.001

Mean differences were obtained from the multivariable linear-regression results reported in the manuscript. The complete coefficients for the remaining model covariates were not available. CI: confidence interval; EPDS: Edinburgh Postnatal Depression Scale; PKR: Pakistani rupees.

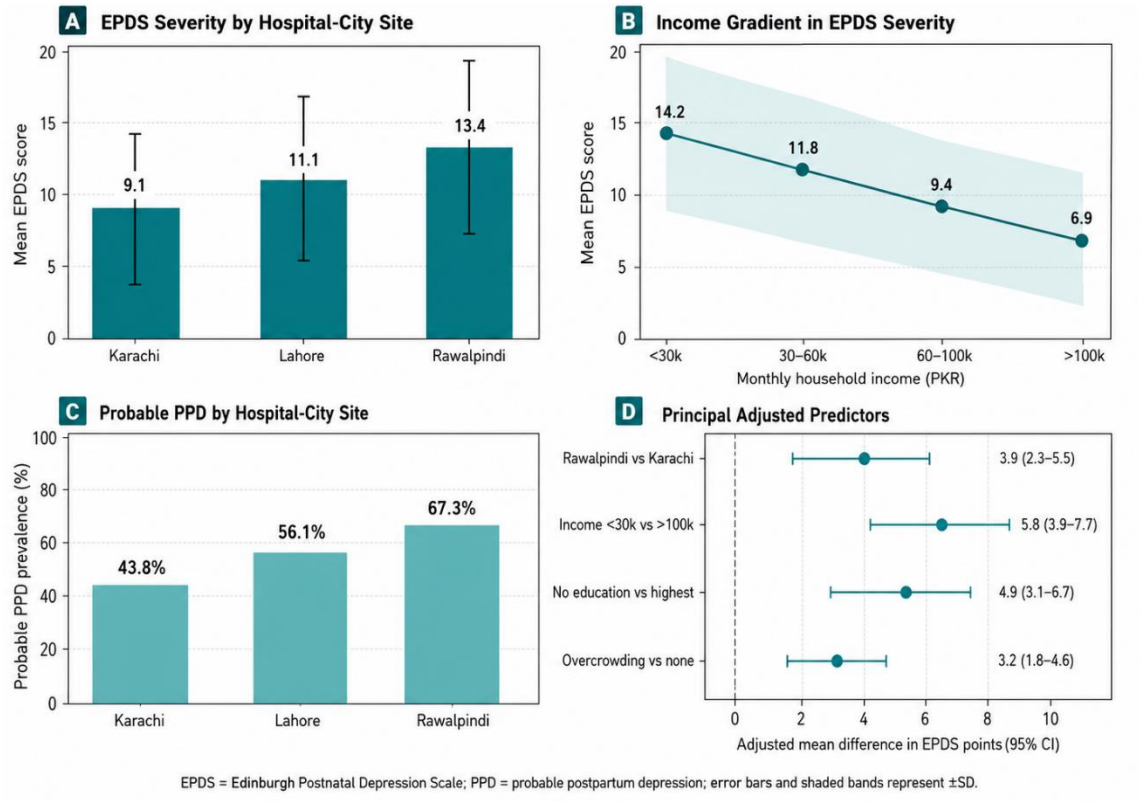

Figure 1: Panel A presents mean EPDS scores with standard deviations across the Karachi, Lahore, and Rawalpindi hospital-city sites. Panel B demonstrates the inverse household-income gradient in mean EPDS severity, with scores decreasing from 14.2 ± 5.5 in households earning less than PKR 30,000 to 6.9 ± 4.6 in those earning more than PKR 100,000. Panel C presents the proportion of women screening positive for probable postpartum depression using an EPDS threshold of ≥10. Panel D displays the principal adjusted predictors of EPDS symptom severity as mean differences with 95% confidence intervals. Rawalpindi had the highest mean EPDS score at 13.4 ± 5.9 and the highest probable postpartum depression prevalence at 67.3%, compared with 9.1 ± 5.2 and 43.8% in Karachi. A consistent socioeconomic gradient was observed, with a 7.3-point

unadjusted difference between the lowest- and highest-income categories. In adjusted analyses, household income below PKR 30,000 showed the largest association with symptom severity, corresponding to a 5.8-point higher EPDS score (95% CI: 3.9–7.7), followed by no formal education at 4.9 points (95% CI: 3.1–6.7), Rawalpindi hospital-city site at 3.9 points (95% CI: 2.3–5.5), and household overcrowding at 3.2 points (95% CI: 1.8–4.6).

The random forest analysis identified household income, maternal education, and hospital-city site as the three leading predictors of EPDS scores. The model achieved a test-set R^2 of 0.51, indicating moderate predictive performance in the held-out sample. Complete numerical variable-importance estimates and test-set RMSE were not reported in the available analytical output and therefore could not be tabulated. The convergence of the regression and random-forest findings nevertheless indicated that socioeconomic disadvantage and hospital-city context were consistently associated with greater postpartum depressive symptom severity within the analysed dataset.

DISCUSSION

This multicentre retrospective record-based study identified consistent hospital-city and socioeconomic differences in postpartum depressive symptom severity among women receiving postnatal care in Karachi, Lahore, and Rawalpindi. Rawalpindi had the highest mean EPDS score and the largest proportion of women screening positive for probable postpartum depression. Low household income, absence of formal maternal education, and household overcrowding were independently associated with higher EPDS scores after adjustment for the available maternal and obstetric characteristics. The random forest analysis produced a test-set R^2 of 0.51 and ranked household income, maternal education, and hospital-city site among the leading predictors, supporting the consistency of the principal patterns across conventional and predictive analytical approaches.

The higher EPDS scores observed at the Rawalpindi hospital-city site should be interpreted as a site-level difference rather than an independent geographic effect. Because one hospital represented each city, the analysis could not distinguish the influence of geographic location from differences in hospital referral patterns, catchment populations, screening practices, socioeconomic composition, or access to postnatal services. Nevertheless, the magnitude of the observed difference was clinically relevant. The mean EPDS score in Rawalpindi was 4.3 points higher than in Karachi in the unadjusted comparison, while the adjusted difference remained 3.9 points. Previous studies from Pakistan and other low-resource settings have similarly reported variation in postpartum depressive symptoms across different clinical and socioeconomic populations, indicating that local service characteristics and social conditions may influence the burden detected through hospital-based screening (1–5).

Household income showed the strongest socioeconomic gradient. Mean EPDS scores declined progressively from 14.2 among women in households earning less than PKR 30,000 per month to 6.9 among those earning more than PKR 100,000. After adjustment, membership in the lowest-income group remained associated with a 5.8-point higher EPDS score. Financial hardship may increase postpartum psychological distress through food insecurity, difficulty meeting healthcare expenses, unstable employment, restricted transport, inadequate living conditions, and dependence on family members. These pressures may be particularly important after childbirth, when household expenditure and caregiving demands increase. The consistency of the income gradient across descriptive, regression, and random-forest analyses strengthens the evidence that economic disadvantage was an important predictive characteristic within the analysed sample, although the retrospective design did not permit a causal interpretation (1,5–8).

Maternal education was also strongly associated with symptom severity. Women with no formal education had a mean EPDS score of 14.8, compared with 9.5 among women in the highest reported education category, while the adjusted difference was 4.9 points. Limited education may affect mental-health literacy, awareness of available services, communication with healthcare professionals, employment opportunities, and autonomy in seeking care. Education may also function as an indicator of broader socioeconomic disadvantage rather than acting as an isolated determinant. Pakistani research has highlighted the importance of social and structural barriers in women's access to mental-health services, particularly when low literacy, stigma, restricted mobility, and limited family support occur together (5–8).

Household overcrowding remained independently associated with a 3.2-point higher EPDS score after adjustment for income, education, and other measured characteristics. This finding suggests that the physical and social living

environment may contribute information beyond household income alone. Overcrowding can reduce privacy, interrupt sleep, intensify domestic responsibilities, and increase interpersonal stress during a period when mothers require rest and emotional support. The mean EPDS difference between overcrowded and non-overcrowded households was 4.5 points in the unadjusted analysis, indicating a potentially important clinical gradient. However, because household composition, caregiving support, room ventilation, noise, and sleep quality were not measured, the mechanisms underlying this association could not be determined.

More than half of the participants screened positive for probable postpartum depression using an EPDS threshold of 10 or higher. This proportion was higher than that reported in several general postpartum populations but may reflect the hospital-based sample, the selected screening threshold, the broad timing of assessment, and differences in clinical or socioeconomic risk. The EPDS is a screening instrument rather than a diagnostic interview; therefore, the reported prevalence represents probable postpartum depression and should not be interpreted as the prevalence of clinically confirmed depressive disorder. The inclusion of assessments conducted between 24 hours and six weeks postpartum may also have increased heterogeneity because transient emotional symptoms in the immediate post-delivery period may differ from persistent depressive symptoms identified later in the postpartum period.

The random forest model explained 51% of the variation in EPDS scores in the held-out test sample. This indicated moderate predictive performance but also showed that a substantial proportion of outcome variation was not captured by the available variables. Important unmeasured predictors may have included antenatal anxiety or depression, previous episodes of depression, pregnancy intention, social support, marital relationship quality, breastfeeding difficulties, neonatal illness, obstetric complications, sleep disturbance, and the exact interval between delivery and EPDS assessment. The ranking of household income, maternal education, and hospital-city site as leading predictors was consistent with the regression findings. However, variable importance reflects predictive contribution within the analysed dataset and does not establish that these characteristics independently caused postpartum depressive symptoms.

This study had several strengths. It included a relatively balanced sample from three tertiary hospitals serving socioeconomically diverse populations. The continuous EPDS score was retained as the primary outcome, preserving more information than a purely binary classification. The use of multivariable regression provided interpretable adjusted mean differences, while random forest regression offered a complementary assessment of predictive importance without requiring all relationships to be specified as linear. Standardised extraction procedures, independent record review, and explicit handling of missing income and education data also strengthened the analytical process.

Several limitations should be considered. The retrospective design depended on the accuracy and completeness of routine clinical documentation and prevented confirmation of the temporal sequence between some exposures and postpartum symptoms. Each city was represented by one hospital, making hospital-level and geographic influences inseparable. The tertiary-care sampling frame may have included women with more complex pregnancies or different socioeconomic characteristics than women receiving primary care, private care, or home-based maternity services. Consequently, the findings should not be generalised to all postpartum women in Pakistan. Women with documented pre-existing serious psychiatric conditions were excluded, which reduced applicability to the wider postpartum population.

The timing of EPDS administration varied from 24 hours to six weeks after delivery and was not reported as a stratified or adjusted predictor. The use of a screening cutoff did not establish a clinical diagnosis. Intimate partner violence was incompletely documented and was excluded from the primary adjusted model because of missing data. Several potentially important psychosocial and obstetric variables were unavailable. The random forest test set was relatively small, and external validation was not performed. The model's predictive performance and variable rankings therefore require confirmation in independent datasets containing multiple hospitals within each geographic region.

The findings have practical implications for postnatal screening. Household income, maternal education, overcrowding, and hospital-site context are characteristics that can be identified during routine maternity care without specialised psychiatric testing. Women experiencing multiple forms of socioeconomic disadvantage may

benefit from earlier EPDS screening, closer clinical assessment, and clear referral pathways. Future prospective studies should use standardised assessment intervals, diagnostic confirmation, repeated measurements, and broader psychosocial data. Sampling multiple hospitals within each city would permit more reliable separation of geographic and institutional effects. Prediction models should also undergo internal resampling, calibration assessment, and independent external validation before clinical implementation.

CONCLUSION

Low household income, absence of formal maternal education, household overcrowding, and the Rawalpindi hospital-city site were associated with greater postpartum depressive symptom severity among women receiving care at three tertiary hospitals in Pakistan. Household income demonstrated the strongest socioeconomic gradient, while the complementary random forest analysis showed moderate predictive performance and identified similar leading predictors. These findings support targeted postpartum mental-health screening among socioeconomically disadvantaged women, but they should be interpreted as hospital-based associations rather than causal or nationally representative geographic effects.

REFERENCES

1. Abdullah MA, Shaikh BT, Sattar NY, Sarwar B, Ahmed AS, Fatima SS. Are social determinants associated with depression among married women of reproductive age? A mixed methods study from urban slums of Islamabad, Pakistan. PLOS Glob Public Health. 2024;4(7):e0003463.

2. Muhammad FS, Shahabudin SM, Talib MBA. Measuring spatial inequalities in maternal and child mortalities in Pakistan: evidence from geographically weighted regression. BMC Public Health. 2024;24(1):2229.

3. Rahman A. Facilitators, barriers and requirements for addressing perinatal depression in urban slums through mobile phone-based interventions in Bangladesh: a qualitative study [dissertation]. Liverpool: University of Liverpool; 2025.

4. Emikpe AO, Obeng H, Bonsu ZM, Tsiko EA, Owusu LB, Dwumfour CK, et al. Predictors of postpartum depression among postnatal women in Kumasi, Ghana. 2025;25(1):1179.

5. Bhatti MMA, Ashraf MS, Bhatti H, Jamil M, Mukhdooma K, Ibrahim Z. Prevalence and social factors associated with postpartum depression in urban areas of Pakistan: a cross-sectional study. J Rawalpindi Med Coll. 2024;28(4).

6. Zafar A, Akram BA, Wasim M, Pires IM, Coelho PJ. PERI_DEP: a dataset of mothers' mental health in Pakistan. Data Brief. 2025;60:111621.

7. Sajjad A, Shah S, Abbas G, Aslam A, Randhawa F, Khurram H, et al. Treatment gap and barriers to access mental healthcare among women with postpartum depression symptoms in Pakistan. 2024;12:e17711.

8. Chachar AS, Mian AI. A review of the intersection of social determinants and child and adolescent mental-health services: a case for social psychiatry in Pakistan. World Soc Psychiatry. 2022;4(2):69–77.

9. Makinde OA, Uthman OA, Mgbachi IC, Ichegbo NK, Sule FA, Olamijuwon EO, et al. Vulnerability in maternal, newborn, and child health in low- and middle-income countries: findings from a scoping review. PLoS One. 2022;17(11):e0276747.

10. Mah B, Brown A, Eades S, Pringle K, Rae K, et al. Psychological distress, stressful life events and social disadvantage in pregnant Indigenous Australian women residing in rural and remote New South Wales: a longitudinal cohort study. 2022;9(6):2197–207.

11. Hirani R, Noruzi K, Khuram H, Hussaini AS, Aifuwa EI, Ely KE, et al. Artificial intelligence and healthcare: a journey through history, present innovations, and future possibilities. 2024;14(5):557.

12. Iqbal F, Satti MI, Irshad A, Shah MA. Predictive analytics in smart healthcare for child mortality prediction using a machine-learning approach. Open Life Sci. 2023;18(1):20220609.

13. Raza SH, Khan L. Beyond numbers: mapping poverty disparities in Pakistan through a spatial lens. 2025;18(1):28.

14. Wuermli A, Hiott MC, Ugarte E, Rahman MS, Elahi M, Rahim A, et al. Cohort profile: a prenatal birth cohort study of intergenerational risk and resilience after conflict and forced displacement. BMJ Open. 2025;15(11):e099052.

15. Sah LK. Social determinants of mental health of pregnant women in Nepal: a sequential exploratory mixed-methods research study [dissertation]. Canterbury: Canterbury Christ Church University; 2024.

16. Holland N. An exploration of the factors that influence seeking treatment from the perspective of Muslim women who have experienced postpartum depression [dissertation]. Cape Town: University of the Western Cape; 2024.

17. Shukla A, Purohit N. Mental health matters: investigating the association of maternal psychological factors with pregnancy outcomes in Central Uttar Pradesh, India. 2026:1–11.

18. Hassan R, Jafree S. Exploring the challenges and coping strategies faced by women street vendors in Central Punjab, Pakistan. Lahore: Department of Sociology, Forman Christian College; 2024.

19. Rahman MM, Sorkar MNU, Nakamura R, Islam MM, Alam MA, Azmat SK, et al. Sociodemographic disparities in health-related quality of life among hypertensive patients in Bangladesh: a comprehensive survey analysis. 2025;34(5):1457–71.

20. Park K. Addressing social isolation and suicide risk in South Korea's permanent rental housing: an empirical investigation of intervention strategies. 2025;11(1).

21. Karim HT. Maternal perceptions of postpartum depression and its influence on children in early years [dissertation]. Dhaka: BRAC University; 2025.

22. Khan T. Household, neighbourhood, and community contexts and the mental health of immigrants and immigrant mothers: an ecological perspective. 2026.

23. Hamidi AR, Ford JD, Novo P, Paavola J, Malik IH. Rethinking flood resilience: systemic risk, governance failures, and the social production of vulnerability in the Himalayan–Indus region. J Disaster Stud Manag. 2026;2(1):8.

24. Arshed M, Kiran M, Ashraf W, Virk MKS, Qamer S, Umer MF, et al. Non-adherence to oral antidiabetic medications among patients with type 2 diabetes: evidence from a nationwide multicentre study in a lower-middle-income country. 2025;25(1):4072.

25. Muqaddam I, Kayani SN, Gul K, Khan M, Murad S, Kanwal A, et al. Effectiveness of community-based health education in reducing home birth rates.

26. Sahni PS, Dewan S, Sachdeva M, Ranathunga N, Fadhullah W, Abdullah JM, et al. Climate cognition crisis: an urgent call to address the effects of air pollution and extreme heat on learning environments for children in South and Southeast Asia. 2025;32(5).