README: Replication Package

****

Title: Predictive Modeling of Women's Mental Health: Financial Confidence as a Humanitarian Determinant in Bangladesh
Conference: 2025 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON ECE), Cox's Bazar, Bangladesh

Authors:
Asma Binte Afzal
Department of Environmental and Occupational Health
Indiana University Bloomington
Indiana, USA
https://orcid.org/0000-0003-4318-376X
Muhammad Hassan Bin Afzal
Department of Statistics and Data Science
Jahangirnagar University
Dhaka, Bangladesh
https://orcid.org/0000-0001-8192-0885

Core research question: To what extent does women's perceived financial confidence in accessing healthcare predict anxiety and depression, and how are these associations shaped by education, digital inclusion, and community context in Bangladesh?

1. Study overview

This README presents the replication materials for the paper "Predictive Modeling of Women's Mental Health: Financial Confidence as a Humanitarian Determinant in Bangladesh," prepared for the 2025 IEEE International Women in Engineering Conference on Electrical and Computer Engineering in Cox's Bazar, Bangladesh.

The study uses data from the Bangladesh Demographic and Health Survey 2022 to examine how women's perceived financial ability to obtain money for medical treatment relates to symptoms of anxiety and depression. The analysis focuses on women aged 15 to 49 and links a financial access measure to the GAD-7 and PHQ-9 mental health scales, while accounting for digital inclusion, socioeconomic status, autonomy, and community context.

2. Related work and BD LIFT index

The BD LIFT Index, which stands for Bangladesh Digital Life Inclusion for Females and Technology, first appears in a separate ACM conference paper that focuses on digital inclusion and women's healthcare decision-making. That paper describes the construction logic and motivation for BD LIFT. In the present WIECON ECE study, BD LIFT serves

as a key covariate, capturing women's digital inclusion context, in analyses examining the link between financial access and mental health.

Reference for the BD LIFT paper:

Muhammad Hassan Bin Afzal. 2025. From Connectivity to Care: How Digital Inclusion Shapes Women's Healthcare Decision-Making in Bangladesh. In the 12th International Conference on Next Generation Computing, Communication, Systems and Security NSysS'25), December 18-20, 2025, Sylhet, Bangladesh. ACM, New York, NY, USA, 10 Pages. https://doi.org/10.1145/3777555.3777557

Replication Package to BD-Lift Study:

Afzal, Muhammad Hassan Bin. 2025. "Replication Package for: From Connectivity to Care: How Digital Inclusion Shapes Women's Healthcare Decision-Making in Bangladesh." Harvard Dataverse. https://doi.org/10.7910/DVN/DSQJU9

3. Data source and ethical compliance

All data originate from the Bangladesh Demographic and Health Survey 2022 Individual Recode file. The authors obtain access through the DHS Program under its standard data use agreement. To respect confidentiality rules and to protect respondents, the replication dataset in this package is an entirely derived analytic file. It does not contain any DHS raw identifiers or reconstructable geographic information.

The shared file removes household numbers, enumeration area codes, global positioning variables, strata labels, and all direct DHS identifiers. Cluster identifiers are replaced with an anonymized community identifier that preserves the multilevel structure needed for modeling and cannot be linked back to the DHS sampling frame. Item-level DHS variables used to build scales and indices do not appear in the public analytic dataset.

4. Research question

The core research question is as follows. To what extent does women's perceived financial confidence in accessing healthcare predict anxiety and depression, and how are these associations shaped by education, digital inclusion, and community context in Bangladesh?

5. Files in this replication package

The Dataverse deposit includes the following core files.

a) mental_health_finconf_custom.dta An anonymized analytic dataset with only derived variables used in the models.

b) Replication_DoFile_WIE2025.A self-contained Stata script that loads the analytic dataset and reproduces the main results.

c) README This document explains data sources, variable construction, modeling choices, and replication steps.

6. Key variables in the analytic dataset

The analytic dataset includes only the variables necessary to reproduce the tables, models, and figures in the paper. Each key variable appears in paragraph form below.

a) gad7_ord4. This ordinal anxiety outcome represents the Generalized Anxiety Disorder scale with four ordered levels. The variable is created from seven GAD-7 items that sum to a total score from zero to twenty-one. The total score is recoded so that 0 represents minimal anxiety, 1 represents mild anxiety, 2 represents moderate anxiety, and 3 represents severe anxiety. This coding follows standard clinical cut points. In the analytic sample, the largest share of women falls into the minimal and mild categories, with fewer cases in the moderate and severe categories.

b) phq9_ord5. This ordinal depression outcome corresponds to the Patient Health Questionnaire scale, with five ordered levels. The variable is created from nine PHQ-9 items that sum to a total score from zero to twenty-seven. The total score is recoded so that 0 represents none to minimal depression, with scores from zero to four, 1 represents mild depression, with scores from five to nine, 2 represents moderate depression with scores from ten to fourteen, 3 represents moderately severe depression with scores from fifteen to nineteen, and 4 represents severe depression with scores from twenty to twenty-seven. In the analytic sample, most women fall in the none to minimal and mild categories, with smaller shares in the higher-severity categories.

c) finconf_med. This binary predictor measures perceived financial access to medical treatment for oneself. It is derived from a BDHS item that asks whether getting the money needed for medical help is a big problem. In this study, the variable is coded as 1 when the respondent reports that money for medical treatment is not a big problem and 0 when the respondent reports that money is a big problem. Conceptually, this variable captures perceived financial access and perceived cost barriers to care rather than a pure personality trait of self-efficacy. Women who report that money is not a big problem experience fewer perceived cost barriers.

d) autonomy_say. This binary variable indicates whether the respondent has a say in her own healthcare decisions. It is derived from a decision-making item that asks who usually makes decisions about the respondent's healthcare. The coding assigns 1 to women who decide alone or jointly and 0 to women who report that someone else chooses. The variable represents health decision autonomy.

e) bdlift. This variable is the BD LIFT index, which captures digital inclusion on a zero-to-three cumulative scale. It combines smartphone ownership, internet use in the last 12 months, and mobile money use for financial transactions. A value of 0 represents no device and no digital activity. A value of 1 represents the device only. A value of 2 represents device plus internet use. A value of 3 represents device, internet, and mobile money use. Higher values indicate stronger digital inclusion and a broader digital environment for women.

f) education. This ordered categorical variable measures educational attainment. It follows the BDHS coding scheme and distinguishes levels such as no education, primary, secondary, and higher education. The models treat education as a factor variable so that comparisons across levels are relative to a reference group. Higher education reflects greater human capital and is expected to be associated with lower mental health risk.

g) wealth. This categorical variable reports the DHS wealth index quintile, ranging from 1 to 5. One represents the poorest quintile, and five represents the wealthiest quintile. The index is constructed by DHS using household assets and living conditions. In the models, wealth is treated as a factor variable capturing relative household socioeconomic position.

h) age. This categorical variable captures age in four ten-year bands. The categories are 15-19, 20-29, 30-39, and 40-49 years. Each category appears as a separate level in the models. This coding allows the analysis to study nonlinear patterns in mental health risk across the reproductive age span.

i) is_working. This binary variable indicates current employment. It equals 1 when the respondent reports that she is currently working for pay or in a recognized occupation and equals 0 when she reports that she is not currently working. Employment status captures one dimension of economic engagement.

j) residence. This categorical variable distinguishes the place of residence. It takes the value 1 for urban residence and 2 for rural residence. The variable reflects contextual differences in access to services and social conditions.

k) bank_account. This binary variable indicates whether the respondent has a bank account in her own name or jointly. It equals 1 if a bank account exists and 0 if there is none. Bank account ownership serves as an additional marker of economic empowerment and a resource for financial access.

l) wt. This continuous variable is the survey weight, constructed by dividing the DHS weight v005 by 1,000,000. It adjusts for unequal selection probabilities and non-response, ensuring that weighted estimates reflect the population of women aged 15 to 49 in Bangladesh in 2022.

m) cluster_id_anon. This identifier represents an anonymized community or cluster. It preserves the multilevel structure of the sample while removing any direct link to the original DHS enumeration areas. The multilevel models include random intercepts at this cluster level.

n) analytic_sample. This binary flag indicates whether women are included in the analytic sample. It equals 1 for respondents who have non-missing values on all outcomes and predictors used in the models and equals 0 otherwise. The replication do file keeps observations with analytic_sample equal to 1.

Table 2: Summary Statistics of Complete-case (N = 15,112). Survey-weighted proportions (95% CI) and weighted means for outcomes
(Estimates from svy: mean on subpopulation cc_all==1; CI = mean ± 1.96×SE)

Role	Variable (BDHS code)	Level	Estimate (95% CI)
Main DVs	GAD-7 ordinal (gad7_ord4)	Minimal (0)	75.30% (74.62, 75.99)
		Mild (1)	21.35% (20.70, 22.01)
		Moderate (2)	2.65% (2.40, 2.91)
		Severe (3)	0.69% (0.56, 0.82)
	PHQ-9 ordinal (phq9_ord5)	None-minimal (0)	72.62% (71.91, 73.33)
		Mild (1)	23.18% (22.51, 23.85)
		Moderate (2)	3.30% (3.02, 3.59)
		Moderately severe (3)	0.75% (0.61, 0.89)
		Severe (4)	0.15% (0.09, 0.21)
	GAD-7 total (gad7_total) — mean	—	**2.94** (2.89, 2.98)
	PHQ-9 total (phq9_total) — mean	—	**3.21** (3.16, 3.26)
Main IVs / Predictors	Financial confidence (finconf_med)	Big problem (0)	47.54% (46.75, 48.34)

		Not a big problem (1)	52.46% (51.66, 53.25)
	Autonomy (autonomy_say)	No say (0)	23.89% (23.21, 24.57)
		Has a say (1)	76.11% (75.43, 76.79)
	BD-LIFT (bdlift)	0 No inclusion	63.00% (62.23, 63.77)
		1 Device only	6.66% (6.27, 7.06)
		2 Device + Internet	15.52% (14.94, 16.09)
		3 Device + Internet + Mobile Money	14.82% (14.25, 15.38)
Controls	Education (v149)	No education	13.59% (13.05, 14.14)
		Incomplete primary	12.65% (12.12, 13.18)
		Complete primary	13.31% (12.77, 13.86)
		Incomplete secondary	33.54% (32.78, 34.29)
		Complete secondary	11.04% (10.54, 11.54)
		Higher	15.87% (15.29, 16.45)
	Residence (v025)	Urban	34.87% (34.11, 35.63)
		Rural	65.13% (64.37, 65.89)
	Bank account (v170)	No	81.17% (80.55, 81.80)
		Yes	18.83% (18.20, 19.45)

	Wealth quintile (v190)	Poorest (1)	17.85% (17.24, 18.46)
		Poorer (2)	19.26% (18.63, 19.89)
		Middle (3)	19.40% (18.76, 20.03)
		Richer (4)	20.37% (19.73, 21.02)
		Richest (5)	23.11% (22.44, 23.79)
	Age (10-yr clusters)	15–19	7.90% (7.47, 8.33)
		20–29	35.10% (34.34, 35.87)
		30–39	33.71% (32.95, 34.46)
		40–49	23.29% (22.61, 23.96)
	Working (is_working)	No	70.29% (69.56, 71.02)
		Yes	29.71% (28.98, 30.44)

7. Analytic sample and survey design

The analytic sample includes women aged 15 to 49 with valid information on GAD-7, PHQ-9, financial confidence, BD LIFT, autonomy, education, wealth, age, employment, residence, bank account status, and sampling weight. The analytic_sample flag identifies these women. The replication do file restricts the data to this group to reproduce the tables and models.

For descriptive statistics, the do file declares the survey design using the anonymized cluster identifier and the weight wt. The command svyset cluster_id_anon pweight wt sets this design. This step allows weighted cross-tabulations of the main variables that represent the target population. The multilevel models below use cluster-level random effects with weights and do not rely on full survey design features, such as strata.

8. Data distributions, correlation, and multicollinearity

The replication do file presents frequency tables for all key variables and summary statistics for the outcomes and weights. The distributions of gad7_ord4 and phq9_ord5 show that most women report none to minimal or mild symptoms, and relatively few women fall into the moderate, moderately severe, or severe categories. This pattern motivates the focus on non-to-minimal versus mild categories in the predictive margins analyses, since these categories cover the bulk of the population and reflect early mental health burden.

The file then computes pairwise correlations among the predictors, including financial confidence, autonomy, BD LIFT, education, wealth, age, employment, residence, and bank account ownership. Correlations remain well below levels that would indicate severe overlap. Variance inflation factors from the collin command are also modest. These results support retaining all predictors in the multivariable models and provide reassurance that multicollinearity does not drive the main findings.

9. Modeling strategy and intraclass correlations

The main inferential approach uses multilevel ordered logistic regression with random cluster-level intercepts. For each outcome, the do file first estimates a bivariate model that includes only finconf_med, and then estimates a complete model that provides for finconf_med, autonomy_say, bdlift, education, wealth, age, employment, residence, and bank_account. These models use the weight wt as a probability weight and treat the cluster_id_anon variable as the random effect.

After fitting the full models, the do file computes intraclass correlation coefficients using the estat icc command. The intraclass correlations indicate that roughly one-eighth of the variation in anxiety and depression lies between communities. This share supports the use of multilevel modeling and shows that community context plays a meaningful role in women's mental health outcomes.

10. Calibration plots

The replication do-file evaluates model calibration for the full GAD-7 and PHQ-9 models. For each outcome, the file computes predicted probabilities from the multilevel ordered logit model, creates deciles of predicted risk, and then calculates the mean predicted probability and the mean observed outcome within each decile. Calibration plots place the observed values against the predicted values across deciles.

These plots allow visual assessment of how closely the model-predicted risk aligns with actual observed severity across the distribution. In this study, the calibration plots show reasonable agreement between predicted and observed values, with no strong pattern of systematic overprediction or underprediction across risk deciles. This agreement supports the adequacy of the multilevel models for describing population patterns.

11. Margins and interaction plots

To aid interpretation, the replication do file uses margins and marginsplot commands to examine how the predicted probability of mild anxiety or mild depression varies by financial confidence, controlling for education, age, wealth, employment, bank account ownership, and BD LIFT levels. The file fits the complete multilevel model for each outcome, then computes predicted probabilities for outcome category 1, which corresponds to mild symptoms.

Interaction plots from marginsplot show the predicted probabilities by categories of education, age, wealth, employment status, bank account ownership, and BD LIFT, separately for women who report that money is a big problem and for those who report that money is not a big problem. The plots show consistent gradients in which higher financial confidence is associated with lower predicted probabilities of mild symptoms across social positions. These margins help translate regression coefficients into more substantive mental health patterns.

12. Proportional odds checks and generalized ordered logit models

Because the outcomes are ordinal, the proportional odds or parallel lines assumption is a key concern. The replication do-file fits single-level ordered logit models for GAD-7 and PHQ-9 that include the same set of predictors as the main models. It then applies the Brant test, with detailed output, to assess whether the coefficients remain constant across thresholds.

The Brant tests indicate that proportional odds hold well for some predictors and are partially violated for others, particularly some secondary covariates. To explore this further, the file estimates generalized ordered logit models using the gologit2 command with the autofit option. This strategy constrains slopes to be equal across thresholds where the data support the proportional odds assumption and relaxes the constraint where the assumption does not hold.

Results from the generalized ordered logit models show that the estimated association between financial confidence and mental health remains stable in direction and size even when some secondary predictors vary across thresholds. This pattern supports the robustness of the main financial confidence effect. It justifies interpreting ordered logit

coefficients as summarizing the overall shift in mental health severity associated with higher financial confidence.

13. Limitations

The study faces several significant limitations. First, the BDHS 2022 data are cross-sectional, so the analysis can describe associations between financial confidence and mental health but cannot establish causal effects. Second, GAD-7 and PHQ-9 rely on self-reported responses, so misreporting and stigma may affect the measured symptoms. Third, some potential confounders, such as exposure to domestic violence and detailed social support measures, are either unavailable in the public DHS file or excluded from the custom dataset to preserve confidentiality.

Lastly, the replication dataset simplifies some elements of the original survey design by removing explicit strata identifiers, and the multilevel models approximate the design using cluster-level random effects and weights. These limitations do not alter the core pattern of findings, but they do call for careful interpretation and for caution in drawing policy conclusions.

14. How to run the replication

Replication in Stata follows simple steps. a) Place the analytic dataset and the replication do file in the same folder. b) Open Stata and change the working directory to that folder. c) Load the dataset using the command use "mental_health_finconf_custom.dta", clear. d) Run the script using the command do "Replication_DoFile_WIE2025.do". The do file then executes all steps in order and reproduces the descriptive tables, multilevel models, intraclass correlations, calibration plots, margins, and proportional odds diagnostics that support the paper.

15. Citation and contact

Users of these replication materials are asked to cite both the WIECON ECE paper and the BD LIFT ACM paper.

WIECON ECE paper. Afzal, Asma Binte, and Afzal, Muhammad Hassan Bin. 2025. Predictive Modeling of Women's Mental Health: Financial Confidence as a Humanitarian Determinant in Bangladesh. Paper prepared for the 11th IEEE International Women in Engineering Conference on Electrical and Computer Engineering, Cox's Bazar, Bangladesh.

BD LIFT paper. Afzal, Muhammad Hassan Bin. 2025. From Connectivity to Care: How Digital Inclusion Shapes Women's Healthcare Decision Making in Bangladesh. Proceedings of the 12th International Conference on Next Generation Computing,

Communication, Systems and Security, NSysS 25, Sylhet, Bangladesh. ACM, New York, NY, USA.

For questions about this replication package, readers may contact Muhammad Hassan Bin Afzal at the Department of Applied Statistics and Data Science, Jahangirnagar University, Dhaka, Bangladesh, or Asma Binte Afzal at the Department of Environmental and Occupational Health, Indiana University Bloomington, Indiana, United States.