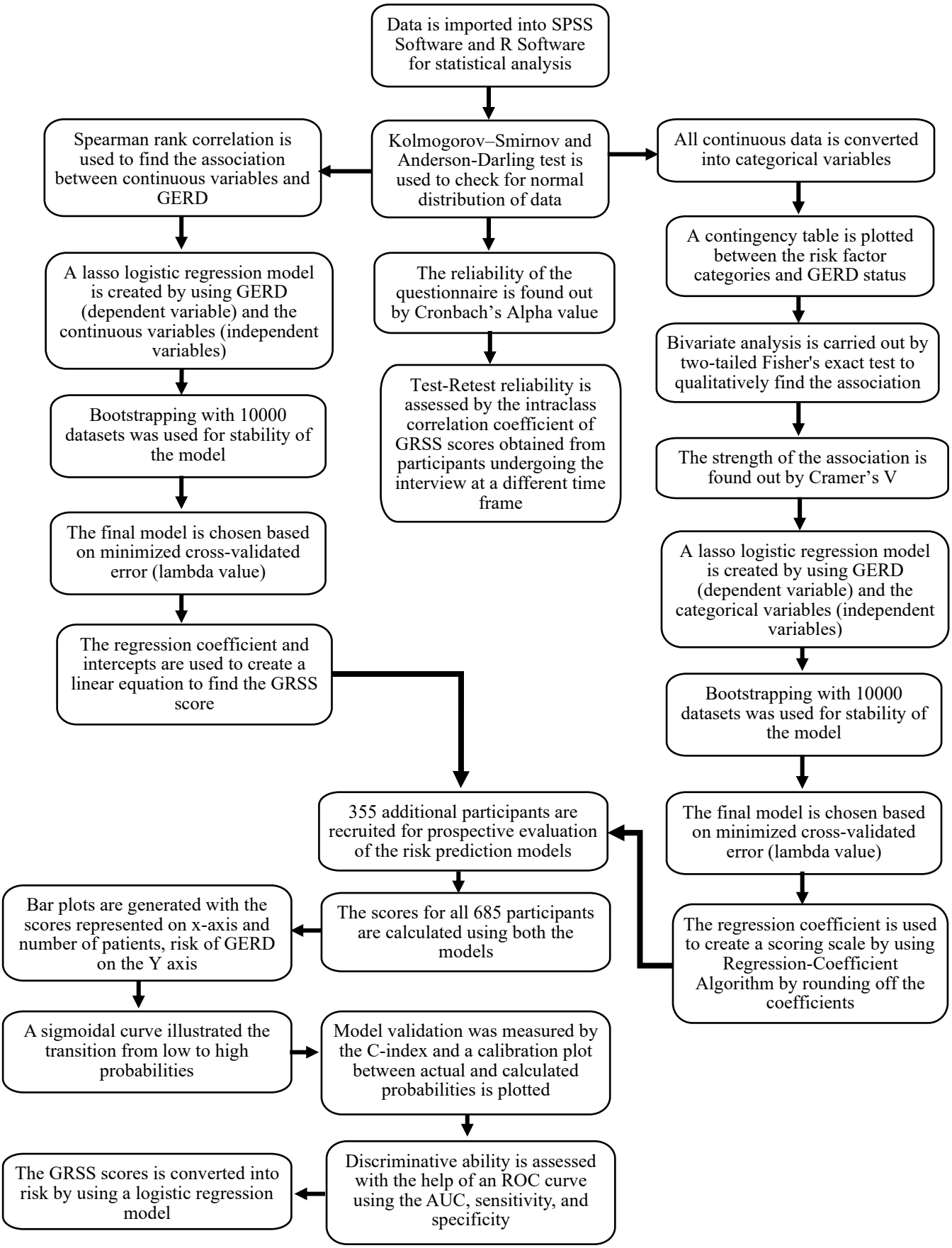
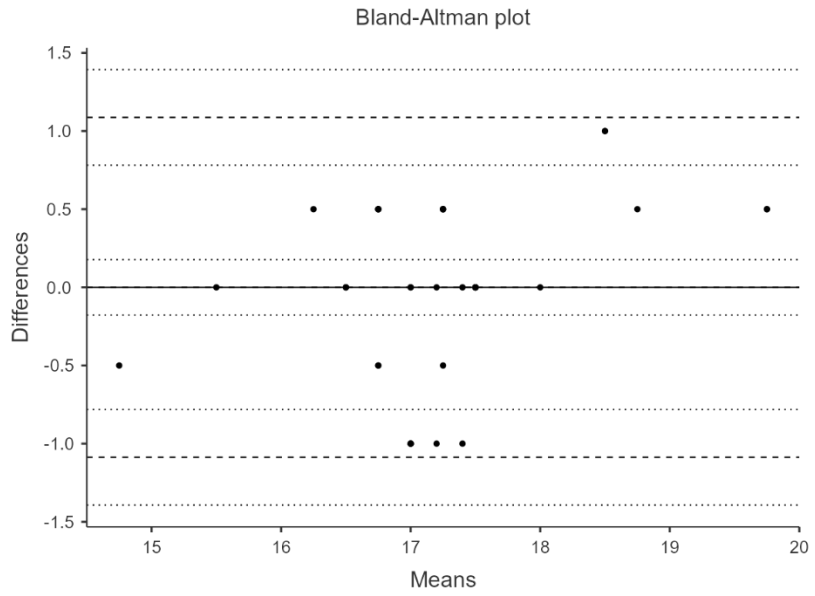
Supplementary material



Supplementary Figure 1 Flowchart illustrating the steps of statistical analysis. AUC: Area under the curve; GERD: Gastroesophageal reflux disease; GRSS: Gastroesophageal Reflux Disease Risk Scoring System; ROC: Receiver operating characteristic; SPSS: Statistical Package for the Social Sciences.



Supplementary Figure 2 Bland-Altmann plot for test-retest reliability.

Supplementary Table 1 Code definitions for categorical variables in the study

Code for age (years)		**Code for diet**		**Codes for alcohol**	
0-20	1	< 4	1	BMI <18	1
20-40	2	4–8	2	BMI 18-22.9	2
40-60	3	8–12	3	BMI 23-24.9	3
60-80	4	> 12	4	BMI >25	4
> 80	5				
Codes for BMI		Codes for stress		Codes for sleep	
Non-alcoholic	1	< 4.5	1	< 4	1
Light drinker (<6)	2	4.5–6	2	4–8	2
Heavy drinker (>6)	3	> 6	3	> 8	3
Codes for smoking		Codes for GERD		Code for community	
Non-smoker	1	Control (GERD-ve)	1	Rural	1

< 5 pack years	2	Case (GERD+ve)	2	Urban	2
> 5 pack years	3				
Codes for socioeconomic class					
Upper class	1				
Upper middle class	2				
Lower middle class	3				
Upper lower class	4				
Lower class	5				

BMI: Body mass index; GERD: Gastroesophageal reflux disease.

Supplementary Table 2 Gastroesophageal Reflux Disease Risk Scoring System Scores risk table

Number	Score	Risk %	Score	Risk %	Score	Risk %	Score	Risk %	Score	Risk %
1	10.5	0.0107	16.9	0.5389	22.9	17.617	28.9	89.4066	34.9	99.7007
6	10.7	0.0121	17	0.5727	23	18.5237	29	89.9729	35	99.7184
11	10.9	0.0137	17.1	0.6087	23.1	19.4661	29.1	90.5122	35.1	99.7351
16	11	0.0146	17.2	0.6469	23.2	20.4445	29.2	91.0254	35.2	99.7508
21	11.1	0.0155	17.3	0.6875	23.3	21.4588	29.3	91.5134	35.3	99.7656
26	11.2	0.0165	17.4	0.7306	23.4	22.5093	29.4	91.9772	35.4	99.7795
31	11.4	0.0186	17.5	0.7764	23.5	23.5958	29.5	92.4178	35.5	99.7926
36	11.5	0.0198	17.6	0.8251	23.6	24.718	29.6	92.836	35.6	99.8049
41	11.6	0.0211	17.7	0.8767	23.7	25.8754	29.7	93.2329	35.7	99.8164
46	11.7	0.0224	17.8	0.9316	23.8	27.0676	29.8	93.6093	35.8	99.8273
51	11.9	0.0253	17.9	0.9899	23.9	28.2937	29.9	93.9661	35.9	99.8376
56	12	0.0269	18	1.0517	24	29.5529	30	94.3042	36	99.8472
61	12.1	0.0286	18.1	1.1174	24.1	30.844	30.1	94.6245	36.1	99.8563
66	12.2	0.0304	18.2	1.1872	24.2	32.1658	30.2	94.9277	36.2	99.8648
71	12.3	0.0324	18.3	1.2612	24.3	33.5167	30.3	95.2146	36.3	99.8728
76	12.4	0.0344	18.4	1.3398	24.4	34.8952	30.4	95.4862	36.4	99.8804
81	12.5	0.0366	18.5	1.4233	24.5	36.2995	30.5	95.7429	36.5	99.8875

86	12.6	0.0389	18.6	1.5118	24.6	37.7275	30.6	95.9857	36.6	99.8941
91	12.7	0.0413	18.7	1.6058	24.7	39.1771	30.7	96.2152	36.7	99.9004
96	12.8	0.0439	18.8	1.7055	24.8	40.6461	30.8	96.4321	36.8	99.9063
101	12.9	0.0467	18.9	1.8113	24.9	42.132	30.9	96.637	36.9	99.9119
106	13	0.0497	19	1.9235	25	43.6323	31	96.8305	37	99.9171
111	13.1	0.0528	19.1	2.0426	25.1	45.1443	31.1	97.0132	37.1	99.9221
116	13.2	0.0561	19.2	2.1688	25.2	46.6654	31.2	97.1857	37.2	99.9267
121	13.3	0.0597	19.3	2.3027	25.3	48.1926	31.3	97.3485	37.3	99.931
126	13.4	0.0635	19.4	2.4446	25.4	49.7233	31.4	97.5021	37.4	99.9351
131	13.5	0.0675	19.5	2.595	25.5	51.2545	31.5	97.647	37.5	99.939
136	13.6	0.0717	19.6	2.7544	25.6	52.7833	31.6	97.7837	37.6	99.9426
141	13.7	0.0762	19.7	2.9233	25.7	54.3069	31.7	97.9127	37.7	99.946
146	13.8	0.0811	19.8	3.1023	25.8	55.8226	31.8	98.0343	37.8	99.9492
151	13.9	0.0862	19.9	3.2918	25.9	57.3274	31.9	98.1489	37.9	99.9522
156	14	0.0916	20	3.4925	26	58.8188	32	98.257	38	99.9551
161	14.1	0.0974	20.1	3.705	26.1	60.2942	32.1	98.3588	38.1	99.9577
166	14.2	0.1035	20.2	3.9298	26.2	61.7511	32.2	98.4548	38.2	99.9603
171	14.3	0.1101	20.3	4.1677	26.3	63.1872	32.3	98.5453	38.3	99.9626
176	14.4	0.117	20.4	4.4194	26.4	64.6002	32.4	98.6306	38.4	99.9648
181	14.5	0.1244	20.5	4.6855	26.5	65.9882	32.5	98.7109	38.5	99.9669
186	14.6	0.1322	20.6	4.9668	26.6	67.3493	32.6	98.7866	38.6	99.9689
191	14.7	0.1406	20.7	5.264	26.7	68.6818	32.7	98.8578	38.7	99.9707
196	14.8	0.1495	20.8	5.578	26.8	69.9842	32.8	98.925	38.8	99.9725
201	14.9	0.1589	20.9	5.9096	26.9	71.255	32.9	98.9882	38.9	99.9741
206	15	0.1689	21	6.2596	27	72.4932	33	99.0478	39	99.9756
211	15.1	0.1796	21.1	6.6288	27.1	73.6978	33.1	99.1038	39.1	99.9771
216	15.2	0.1909	21.2	7.0182	27.2	74.8678	33.2	99.1566	39.2	99.9785
221	15.3	0.2029	21.3	7.4286	27.3	76.0028	33.3	99.2063	39.3	99.9797
226	15.4	0.2157	21.4	7.861	27.4	77.1022	33.4	99.2532	39.4	99.9809
231	15.5	0.2293	21.5	8.3163	27.5	78.1657	33.5	99.2972	39.5	99.9821
236	15.6	0.2437	21.6	8.7954	27.6	79.1932	33.6	99.3387	39.6	99.9831

241	15.7	0.2591	21.7	9.2994	27.7	80.1845	33.7	99.3778	39.7	99.9841
246	15.8	0.2754	21.8	9.8291	27.8	81.1399	33.8	99.4145	39.8	99.9851
251	15.9	0.2928	21.9	10.3856	27.9	82.0595	33.9	99.4491	39.9	99.986
256	16	0.3112	22	10.9697	28	82.9437	34	99.4817	40.1	99.9876
261	16.1	0.3308	22.1	11.5824	28.1	83.7929	34.1	99.5123	40.3	99.989
266	16.2	0.3516	22.2	12.2247	28.2	84.6077	34.2	99.5412	40.4	99.9897
271	16.3	0.3738	22.3	12.8973	28.3	85.3886	34.3	99.5683	40.6	99.9909
276	16.4	0.3973	22.4	13.6013	28.4	86.1365	34.4	99.5939	40.8	99.9919
281	16.5	0.4223	22.5	14.3373	28.5	86.8519	34.5	99.6179	40.9	99.9924
286	16.6	0.4488	22.6	15.1062	28.6	87.5358	34.6	99.6405	41.1	99.9933
291	16.7	0.4771	22.7	15.9087	28.7	88.1889	34.7	99.6618	41.3	99.994
296	16.8	0.507	22.8	16.7455	28.8	88.8122	34.8	99.6818		

Sample Size Calculation

107 Cases Patients + 214 Controls

The sample size was calculated using the formula by Kelsey et. al.

$$n1 = \frac{\left(Z_{\frac{\alpha}{2}} + Z_{1-\beta}\right)^2 \bar{p}\bar{q}(r+1)}{r(p_1 - p_2)^2}$$

$$n2 = r * n1$$

where n1 is the number of cases, n2 is the number of controls, $Z_{\frac{\alpha}{2}}$ is the standard normal deviate for the two-tailed test based on the alpha level (relates to the confidence interval level), Z_β is the standard normal deviate for the one-tailed test based on beta level (relates to the power level), r is the ratio of control to cases, p_1 is the proportion of cases with exposure, p_2 is the proportion of controls with exposure

$q_1 = 1 - p_1$, $q_2 = 1 - p_2$, $\bar{p} = \frac{p_1 + rp_2}{r+1}$, $\bar{q} = 1 - \bar{p}$

Two-sided confidence level (1-alpha) : 95

Power (% chance of detecting) : 80

The ratio of Controls to Cases : 2

Hypothetical proportion of controls with exposure : 53

Hypothetical proportion of cases with exposure	:	71.6
The least extreme Odds Ratio to be detected	:	2.24

The sample size is rounded off to the nearest ten. The sample size for the study is 110 cases and 220 controls.

GRSS QUESTIONNAIRE

Date of Interview:

I) GENERAL DETAILS

Name: Age: Address: OP/Ward Number: Sex:

Weight: Contact number:

II) SOCIO-ECONOMIC STATUS

1. Total monthly income of the family (₹):
2. Head of family – Education: Occupation:
3. No. of family members:

III) SMOKING HISTORY

1. Do you have a habit of smoking?

□ Yes □ No

If yes, answer the following questions or skip to the next section IV

2. What do you smoke?

□ Cigarette □ Beedi □ Cigars □ e-Cigarette □ Cigars □ Pipes

3. How many pieces do you smoke on an average day? _____
4. How many years have you been smoking? _____
5. Have you tried to quit smoking?

□ Yes □ No

IV) ALCOHOL HISTORY

1. Do you have a habit of consuming alcohol?

□ Yes □ No

If yes, answer the following questions or skip to the next section V

2. What type of alcoholic beverage do you consume?

□ Whiskey □ Brandy □ Wine □ Beer □ Spirit □ Vodka □ Rum

3. How much do you consume per day?

□ Quarter □ Half □ Full □ If you consume more, mention _____

4. How frequently do you consume alcohol?

□ Less than 1 time monthly □ 8 to 12 times a month

□ 2 to 4 times a month □ More than 16 times a month

5. Have you ingested over 750 milliliters (equivalent to three-quarters) of alcohol during a single occasion?

□ Never □ Less than once a month □ Approximately once a month □ Weekly □ Daily

6. How many glasses of alcohol do you consume typically in one sitting?

□ 1 - 2 glasses □ 3 - 4 glasses

□ 5 - 6 glasses □ 7 - 9 glasses

□ More than 10 glasses

7. Has anyone expressed concern about your excessive drinking habits?

□ No □ Yes, more than a year ago □ Yes, in the last year

V) SLEEP HISTORY

1. What is your usual bedtime on an average day? _____
2. What time do you wake up in the morning? _____

3. How many hours of uninterrupted sleep do you usually get at night? _____

4. How much time does it usually take for you to fall asleep? □ <30 minutes □ >30 minutes

5. Have you consumed medication to sleep in the past month?

□ Never □ 4 to 8 times a month

□ Less than 4 times a month □ More than 8 times a month

6. Rate your sleep quality using a scale from 1 to 4, where 1 represents excellent and 4 corresponds to poor sleep quality. _____

7. How frequently have you taken more than 30 minutes to fall asleep?

□ Never □ 4 to 8 times a month

□ Less than 4 times a month □ More than 8 times a month

VI) DIET HISTORY

1. How many meals do you consume on an average day? _____
2. At what time do you take the last meal of your day? _____
3. How often do you eat Junk food?

   □ Less than one time a week □ One to Two times a week

   □ Three to Four times a week □ More than Five times a week
4. What type of Diet do you consume?

□ Mixed □ Pure Non-Vegetarian □ Vegetarian □ Keto Diet □ ~~Any Other~~
5. How many portions of fruit do you eat on average per day?

   □ 0 Portions □ 2 Portions

   □ 1 Portion □ ≥ 3 Portions
6. How many serving spoons of vegetables do you eat on average per day?

   _____spoons
7. How many days per week on average do you eat wheat products

   _____ days
8. If there are two options for dinner, what would you choose?

   □ White Rice, Idly, Dosa, Potato Curry □ Green Peas, Ragi Kali, Roti,

Chapati,

9. How many days per week do you include legumes like peas, lentils, and pulses in your diet

_____ days

10. How many days per week on average do you eat spicy foods?

_____ days

11. On a scale of 1 to 5, How spicy do you prefer your food to be? _____

12. How many days do you eat more than the average amount of food per day?

_____ days

VII) STRESS LEVELS

Rate your responses to the following questions on a scale of 1 to 5, where 1 indicates that you have never experienced it, and 5 indicates that you have felt it very often.

1. How many times have you experienced a lack of control over crucial events in your life?_____

2. How often did you feel self-assured in managing personal challenges in your life? _____

3. How frequently did you perceive that your life events unfolded as you had envisioned? _____

4. How many times have you felt that you can't handle your life's problems? _____

5. Do you indulge in any de-stressing activities?

□ Yes □ No

6. Is the destressing activity helpful?

□ Yes □ No

7. What is the effect of your stress on food consumption?

□ Positive □ Negative □ No Effect

Confirmatory factor analysis

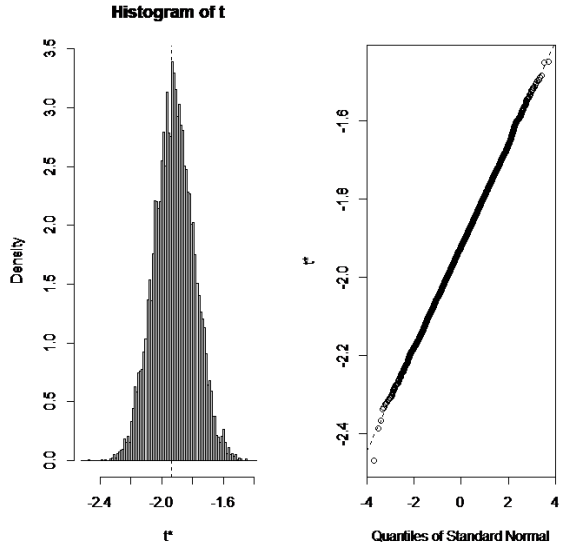
Confirmatory factor analysis was conducted for each Gastroesophageal Reflux Disease Risk Scoring System (GRSS) module to assess the underlying latent constructs. For the Sleep module, items such as "how often do you have heartburn" (standardized loading = 0.566, $P < 0.001$), "how often do you wake up due to discomfort" (loading = 0.764, $P < 0.001$), and "how often do you have difficulty falling asleep" (loading = 0.410, $P < 0.001$). In the Diet module, items such as "how often do you eat oily food" (loading = 0.794, $P < 0.001$), "how often do you skip meals" (loading = 0.477, $P < 0.001$), and "how frequently do you consume fried foods" (loading = 0.676, $P < 0.001$) demonstrated strong and significant loadings. In the stress module, strong loadings were observed for "how often do you feel you can't handle life's problems" (loading = 0.851, $P < 0.001$), "how often do you experience a lack of control over events" (loading = 0.752, $P < 0.001$), and "how often do you feel that life events unfold as envisioned" (loading = 0.313, $P < 0.001$, inverse loading).

Across all modules, items demonstrated statistically significant associations with their respective latent factors ($P < 0.05$), reinforcing the theoretical structure of the GRSS questionnaire. The confirmatory factor analysis models for each GRSS module demonstrated acceptable construct validity, with all items showing statistically significant standardized loadings ($P < 0.05$). For instance, the Sleep module yielded a $\chi^2/df = 2.83$, Comparative Fit Index (CFI) = 0.821, Tucker-Lewis Index (TLI) = 0.701, root mean squared error of approximation (RMSEA) = 0.093, and standardized root mean square residual (SRMR) = 0.083, indicating a reasonable approximation to model fit. The Diet module showed a $\chi^2/df = 2.19$, CFI = 0.921, TLI = 0.841, RMSEA = 0.079, and SRMR = 0.063, consistent with good item-level performance and adequate fit. Similarly, the Stress module yielded $\chi^2/df = 2.92$, CFI = 0.909, TLI = 0.817, RMSEA = 0.095, and SRMR = 0.081. Despite some indices falling slightly below conventional thresholds, the models demonstrated coherent latent structures and strong factor loadings across all domains. The presence of statistically significant loadings and adequate SRMR values suggest a coherent underlying construct, consistent with the theoretical expectations of the GRSS.

Logistic least absolute shrinkage and selection operator regression analysis and

Gastroesophageal Reflux Disease Risk Scoring System model development

Approach 1–using continous variables



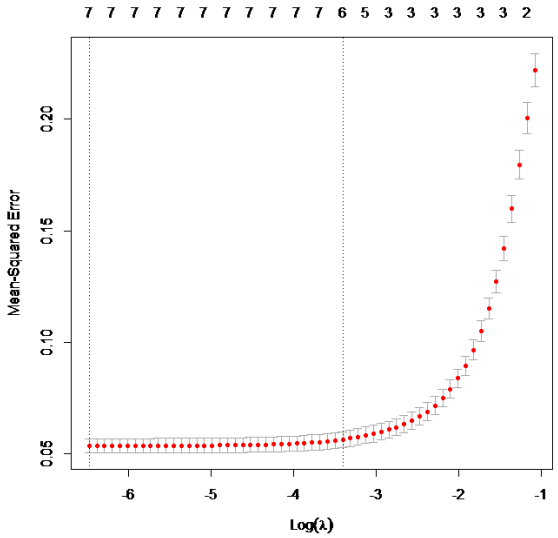


Figure A: Variability of coefficients across bootstrap samples

This figure illustrates the variability of the logistic lasso regression coefficients across 10000 bootstrap samples. Each line represents a different variable, showcasing the stability and variability of the coefficients under the optimal regularization parameter (lambda = 0.95).

figure B: Regularization paths of coefficients

This figure demonstrates the regularization paths of the coefficients as the penalty parameter varies. It shows how different coefficients shrink towards zero with increasing regularization, indicating the impact of the lasso penalty on model complexity and feature selection.

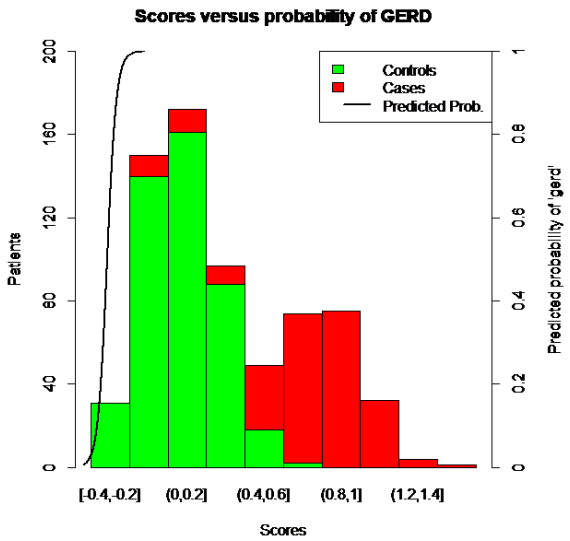


Figure C: Relationship between Gastroesophageal Reflux Disease Risk Scoring System (GRSS) Score and predicted gastroesophageal reflux disease (GERD) probability This sigmoidal curve represents the relationship between the GRSS score and the predicted probability of GERD. The curve shows the transition from low to high probabilities, illustrating how the GRSS score can be used to estimate GERD risk.

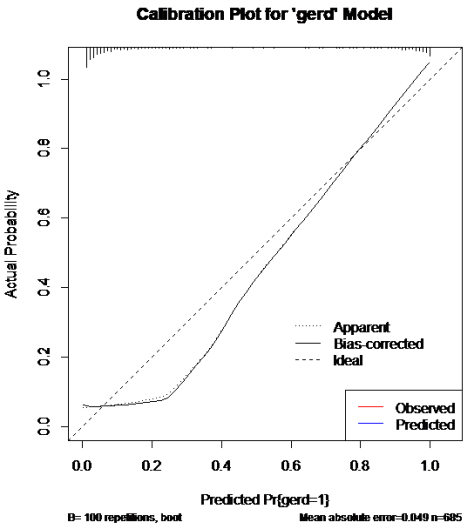


Figure D: Calibration plot of observed *vs* predicted probabilities The calibration plot visually represents the agreement between the observed and predicted probabilities of GERD. Points close to the diagonal line indicate good calibration, where predicted probabilities closely match the actual outcomes.

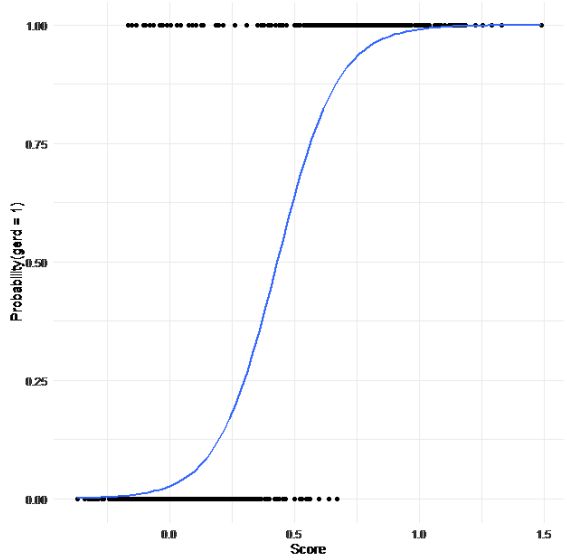


Figure E: Visual representation of GERD risk conversion

This figure visually represents the conversion of GRSS scores into the risk of GERD. The plot demonstrates how the computed GRSS score can be translated into a quantifiable risk of developing GERD, facilitating practical use in clinical settings.

Approach 2–using categorical variables

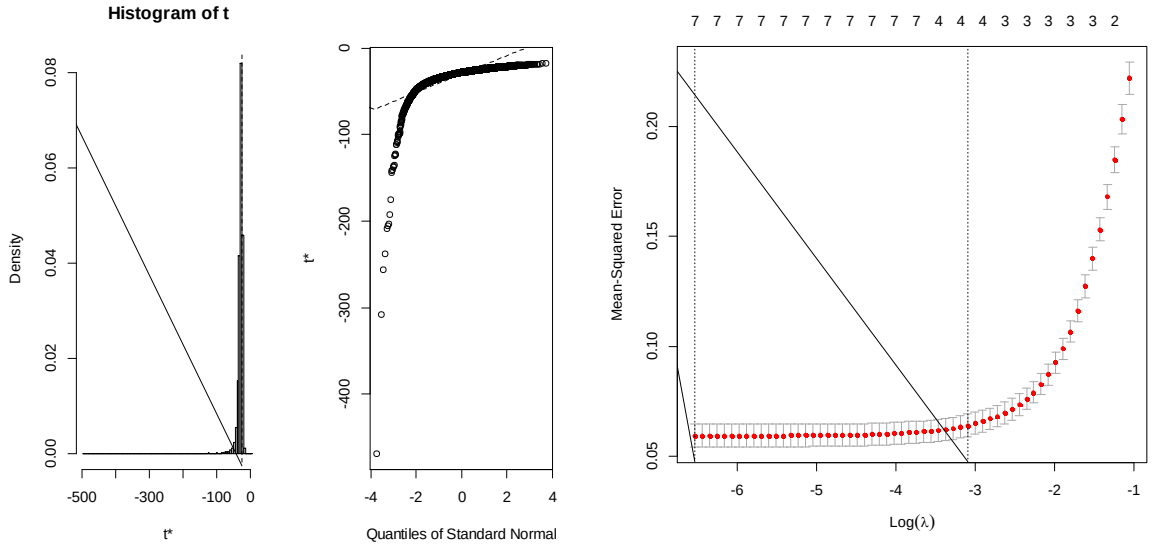


Figure F: Coefficient variability across bootstrap samples
This figure demonstrates the variability of logistic lasso regression coefficients across 10000 bootstrap samples, showing how the coefficients fluctuate and stabilize, contributing to the reliability and robustness of the model.

Figure G: Regularization paths
This plot illustrates the regularization paths for the logistic lasso regression model, showing how coefficients evolve with varying lambda during cross-validation. It highlights the impact of different penalty parameters on the model's coefficients.

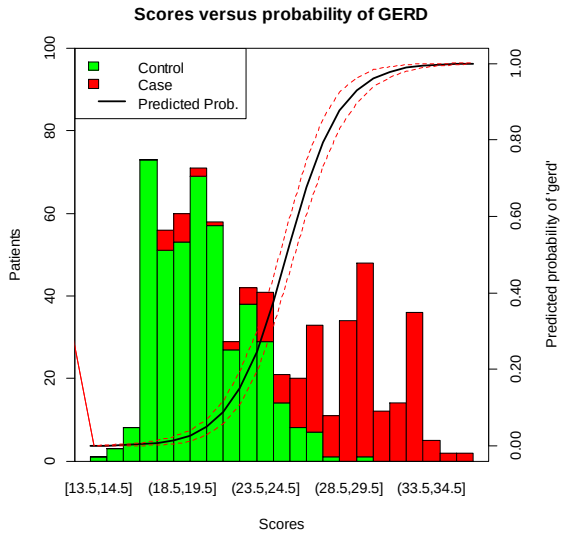


Figure H: Relationship between GRSS score and predicted probability of GERD
This figure depicts the relationship between the GRSS score and the predicted probability of GERD. The scatter plot shows the distribution of patient scores, while the overlaid sigmoidal curve illustrates the transition from low to high GERD probability as the GRSS score increases.

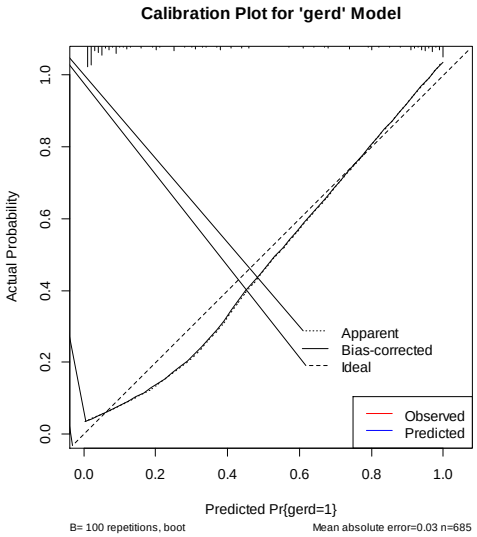


Figure I: Calibration plot
The calibration plot visually represents the agreement between observed and predicted probabilities of GERD. The diagonal line indicates perfect calibration, and the closer the plotted points are to this line, the better the model's predictive accuracy.

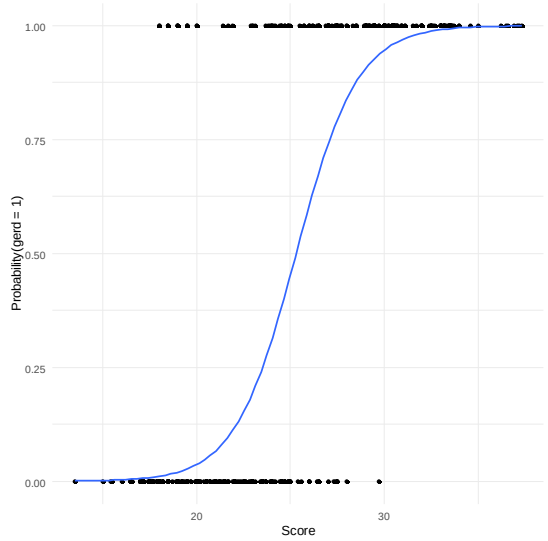


Figure J: Conversion of GRSS Scores to GERD risk

This figure provides a visual representation of how GRSS scores can be converted into the risk of GERD using the developed equation. It helps in understanding the practical application of the scoring scale in predicting GERD risk.