APPENDIX S1

Methods of Elastic Net Logistic Regression

As a complementary multivariable classification analysis, we fitted an elastic net penalized logistic regression model to distinguish patients from healthy controls. The binary outcome was group status. Candidate predictors included the scale-based variables CTQ-Total, DES-28, BDI, BAI, BSI-44, TAS-Total, WI_7, and SSAS, while sex, age, and education duration were treated as forced covariates and retained in all models. This specification was chosen to preserve adjustment for basic demographic factors while allowing the scale variables to compete within a regularized multivariable framework.

The rationale for using elastic net was both statistical and substantive. First, several candidate scale variables represented related psychological and psychosomatic domains and were therefore expected to be moderately intercorrelated. Under such conditions, ordinary multivariable logistic regression may produce unstable coefficients and inflated variance estimates, especially in a modest sample. Elastic net regularization was preferred because it combines shrinkage with variable selection and, unlike pure LASSO, is better suited to settings in which correlated predictors may carry overlapping but still meaningful information. In particular, the elastic net has a grouping property, such that correlated predictors tend to enter or leave the model together rather than in an arbitrarily unstable manner.

To implement this structure, the demographic covariates (sex, age, education duration) were assigned a penalty factor of zero, ensuring that they remained in the model across all iterations, whereas the scale variables were assigned a penalty factor of one and were therefore subject to shrinkage and possible exclusion. This design reflects the analytic goal of estimating the incremental discriminative contribution of the scale variables after demographic adjustment, rather than allowing demographic covariates to be removed by the penalization procedure. Accordingly, variable-selection summaries reported for this model refer only to the scale predictors.

Model tuning and internal validation were performed within a nested cross-validation framework. In the outer loop, the dataset was partitioned into five approximately stratified folds so that the case-control balance was preserved as closely as possible across folds. In each iteration, one fold served as the test set and the remaining four folds served as the training set. Within each outer training set, an inner five-fold cross-validation procedure was used to optimize the elastic net mixing parameter (alpha) over a grid ranging from 0.1 to 0.9. For each candidate alpha, the corresponding regularization path was estimated and the final penalty parameter (lambda) was selected according to the lambda.1se rule, favoring a more parsimonious and conservative solution.

Nested cross-validation was chosen deliberately to reduce optimistic bias in performance estimation. In limited samples, ordinary k-fold cross-validation can yield overly favorable estimates because model development and evaluation are insufficiently separated. Simulation work by Vabalas et al. (1) showed that this bias can remain substantial even at much larger sample sizes, whereas nested cross-validation provides more robust and less biased internal validation when tuning and model selection are part of the analytic pipeline. For this reason, nested resampling was considered more appropriate than single-layer cross-validation for the present predictive analysis.

Within each outer fold, predicted probabilities from the fitted inner-loop–optimized model were generated for both the training and test data. The classification threshold was determined from the training portion using the Youden index, and this threshold was then applied to the corresponding held-out outer test fold. For each outer fold, we calculated the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, positive predictive value, negative predictive value, balanced accuracy, and F1 score. Mean values across the five outer folds were used to summarize overall model performance.

To assess selection stability among the penalized predictors, we additionally recorded how often each scale variable was retained across the outer folds. This was done because, in modest samples with correlated predictors, a variable that is selected consistently across resamples may be more informative than one that appears only sporadically. After completion of nested cross-validation, the model was refit on the full dataset using the same tuning strategy, and the final coefficients were extracted. These final coefficients should be interpreted as penalized estimates, reflecting shrinkage and regularized variable selection, rather than as conventional maximum-likelihood logistic regression coefficients. Consequently, the corresponding odds ratios are best understood as relative contributions within the penalized model rather than as inferential effect estimates accompanied by standard Wald-based confidence intervals.

Results of Elastic Net Logistic Regression

The elastic net model with demographic covariates forced showed moderate discriminative performance. Across the five outer folds, the mean selected alpha was 0.58, the mean AUC was 0.780 (SD = 0.066), mean accuracy was 0.700, mean sensitivity was 0.653, mean specificity was 0.747, mean balanced accuracy was 0.700, and mean F1 score was 0.683. This pattern suggests a somewhat better performance in correctly identifying controls than cases, as specificity exceeded sensitivity. Fold-specific AUC values ranged from 0.667 to 0.840, indicating moderate variability but consistently above-chance discrimination across folds. The number of selected scale predictors ranged from 5 to 7 across outer folds.

Selection-frequency analyses showed that DES-28, BDI, BSI-44, WI-7, and SSAS were retained in all five folds (selection rate = 1.00), identifying them as the most stable penalized predictors. TAS_Total was retained in four of five folds (selection rate = 0.80), whereas BAI was retained in only two folds (selection rate = 0.40). CTQ-Total was not selected in any fold (selection rate = 0.00). This pattern suggests that dissociation, depressive symptoms, somatic symptom burden, health anxiety, and somatic amplification contributed the most stable discriminative information beyond the forced demographic covariates, whereas childhood trauma severity did not demonstrate incremental value in this multivariable penalized framework.

In the final full-data refit, the selected alpha was 0.4. As specified, the forced demographic covariates remained in the model: male sex (beta = 0.395, OR = 1.484), age (beta = −0.014, OR = 0.986), and education duration (beta = −0.095, OR = 0.909). Among the scale variables, the final retained predictors were DES-28 (beta = −0.021, OR = 0.979), BDI (beta = 0.040, OR = 1.041), BSI-44 (beta = 0.014, OR = 1.014), TAS-Total (beta = 0.013, OR = 1.013), WI-7 (beta = 0.105, OR = 1.111), and SSAS (beta = 0.040, OR = 1.041). By contrast, CTQ-Total and BAI were shrunk to zero and excluded from the final penalized component of the model. Among the retained scales, WI-7 showed the largest odds ratio, suggesting the strongest relative contribution within the final elastic net solution.

Overall, these findings indicate that, after adjustment for sex, age, and education, the most stable multivariable discriminators of chronic pruritus without a primary dermatosis or identifiable systemic disease were health anxiety, depressive symptoms, somatic symptom burden, somatic amplification, alexithymia, and dissociation, whereas childhood trauma severity and, in the final model, anxiety symptom severity did not show comparable independent discriminative value. These coefficients should be interpreted as penalized estimates reflecting shrinkage and variable selection rather than conventional maximum-likelihood logistic regression coefficients.

Multivariable logistic regression based on elastic net–selected predictors

To complement the penalized analysis and facilitate conventional interpretability, an unpenalized multivariable logistic regression model was fitted including the predictors retained by the elastic net together with the prespecified covariates (age, sex, and education). This approach was used to obtain interpretable effect estimates (odds ratios and 95% confidence intervals) while preserving the variable-selection structure identified by the penalized model.

In this model, DES-28 (OR = 0.907, p = .001), BDI (OR = 1.105, p = .004), TAS-Total (OR = 1.047, p = .036), and SSAS (OR = 1.091, p = .019) were significantly associated with case status. Sex was also significant (OR = 0.318, p = .030), indicating lower odds for the coded category relative to the reference group. Age, education, BSI-44, and WI-7 were not statistically significant in the multivariable model ($p > .05$).

The overall model was significant ($\chi^2(9) = 64.45$, $p < .001$), explaining a moderate proportion of variance (Nagelkerke $R^2 = .466$) and showing an overall classification accuracy of 77.3%. The pattern of results was broadly consistent with the elastic net findings, with the same core set of predictors contributing to case–control discrimination. Full model coefficients and additional diagnostics are presented in Tables SX–SXI.

1. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. PLoS One 2019; 14: e0224365. https://doi.org/10.1371/journal.pone.0224365.