Appendix 1

Short note:

This appendix provides detailed technical information on the sample size calculation, blood pressure averaging, and model fit indices that were summarized in the main Methods section. These details are included here to ensure transparency and reproducibility while maintaining the readability of the primary manuscript. Additionally, formulas removed from the Methods section in response to Reviewer 1's feedback are reinstated here.

Sample Size Calculation

The sample size for this study was determined using a power analysis to ensure sufficient statistical power for detecting associations in a multistage cluster design. The calculation accounted for the following parameters:

- **Confidence Level**: 95% (corresponding to (z = 1.96))
- **Expected Proportion (p)**: 0.5 (maximizing variance for sample size estimation)
- **Precision (d)**: 0.05
- **Design Effect (DEFF)**: 1.5 (to account for the multistage cluster sampling design)
- **Non-response Rate (NRR)**: 0.1 (10%)

The sample size was calculated using the formula:

$$n = \frac{d^2 Z^2_{1-\alpha/2} \times p(1-p) \times \text{DEFF} \times (1-\text{NRR})}{1}$$

For calculating the cluster effect, the following equation was used:

$$\text{DEFF} = 1 + (m-1)\rho$$

This yielded a required sample size of approximately 200 participants, ensuring adequate power for the study's objectives.

Blood Pressure Averaging

Blood pressure was measured using a calibrated Omron M7 Intelli IT digital sphygmomanometer. Three consecutive readings were taken at five-minute intervals, and the final blood pressure value was calculated as the average of these readings to ensure accuracy. The formula for the final blood pressure value is:

$$\text{BP}_{\text{final}} = \frac{3}{\sum_{i=1}^{3} \text{BP}_i} \pm 2\sqrt{\sum_{i=1}^{3} (\text{BP}_i - \overline{\text{BP}})^2}$$

This approach minimized measurement error and provided a reliable classification of prehypertension (systolic: 120–139 mmHg, diastolic: 80–89 mmHg).

Model Fit Indices for Structural Equation Modeling (SEM)

The Health Belief Model (HBM) was validated using Structural Equation Modeling (SEM) to assess the relationships between its constructs and blood pressure control behaviors. The following model fit indices were used to evaluate the SEM model's goodness of fit:

- **Comparative Fit Index (CFI)**: 0.942 (threshold: >0.90 for acceptable fit, >0.95 for good fit)
- **Tucker-Lewis Index (TLI)**: 0.938 (threshold: >0.90 for acceptable fit, >0.95 for good fit)
- **Root Mean Square Error of Approximation (RMSEA)**: 0.048 (95% CI: 0.042–0.054) (threshold: <0.08 for acceptable fit, <0.05 for good fit)
- **Standardized Root Mean Square Residual (SRMR)**: 0.039 (threshold: <0.08 for good fit)

These indices confirmed that the SEM model provided a robust fit to the data, supporting the validity of the HBM constructs in predicting blood pressure control behaviors.

Machine Learning Model Configurations

The machine learning (ML) models used in this study—Random Forest, Support Vector Machine (SVM), Gradient Boosting, and Neural Networks—were configured and optimized as follows:

- **Random Forest**: Implemented with 100 trees, using k-fold cross-validation (k=5) to prevent overfitting. Feature importance was assessed using Gini impurity.
- **Support Vector Machine (SVM)**: Employed with a radial basis function (RBF) kernel. Hyperparameters (C and gamma) were tuned using grid search.
- **Gradient Boosting**: Configured with 200 estimators, a learning rate of 0.1, and maximum depth of 3. Hyperparameters were optimized via grid search.
- **Neural Networks**: A deep learning architecture with two hidden layers (64 and 32 neurons, respectively), using ReLU activation and dropout (0.3) to prevent overfitting. The model was trained over 100 epochs with early stopping based on validation loss.

Model performance was evaluated using the following metrics:

- Sensitivity
- Specificity
- Area under the ROC curve (AUC)

These configurations ensured that the ML models were optimized for predicting blood pressure control behaviors, with the Gradient Boosting model achieving the highest AUC of 0.895.