



# Predicting Adherence to Complex Health Regimens in Women with Chronic Illness: A Random Forest Approach Integrating Health Literacy and Self-Efficacy


Abulhudha. Subair<sup>1</sup>, Aditya. Prasetyo<sup>2\*</sup>

<sup>1</sup> Faculty of Health, Universitas Kurnia Jaya Persada Palopo, Indonesia



<sup>2</sup> Department of Educational Psychology, Universitas Gadjah Mada, Yogyakarta, Indonesia

\* Corresponding author email address: aditya.prasetyo@ugm.ac.id

### Editor

Parvaneh Mohammadkhani<sup>1</sup>  
Professor, Department of  
Psychology, University of Welfare  
and Rehabilitation Sciences,  
Tehran, Iran  
p.mohammadkhani@uswr.ac.ir

### Reviewers

**Reviewer 1:** Merve Okan<sup>1</sup>  
Department of Psychology, University of Toronto, Toronto, Canada.  
Email: merveokan@mail.utoronto.ca  
**Reviewer 2:** Kacper Jankowski<sup>1</sup>  
School of Psychology, Australian Catholic University, Brisbane, Queensland,  
Australia. kacper.jankowski@acu.edu.au

## 1. Round 1

### 1.1. Reviewer 1

Reviewer:

The Methods section provides a generally clear description of the multicenter, cross-sectional design and inclusion/exclusion criteria, but the use of purposive sampling across four referral hospitals raises concerns about selection bias and limits to external validity; these implications for generalizability should be explicitly discussed, and the authors should clarify whether any recruitment flow diagram or comparison with the broader clinic population is available to contextualize sample representativeness.

The measurement strategy is a strength in that it uses established scales for adherence, health literacy, and chronic disease self-efficacy; nevertheless, the manuscript should provide more detail on the psychometric properties of these instruments specifically in the current Indonesian female chronic-illness population (e.g., prior local validation studies, factor structure, internal consistency coefficients from the present sample, and any cultural or linguistic adaptations made), as this is critical for interpreting both descriptive and predictive findings.

The choice of Random Forest as the primary analytic method is appropriate for modeling non-linear, multidimensional relationships and handling potential multicollinearity; however, the description of the machine-learning pipeline is not

sufficiently detailed for reproducibility—please report the exact data preprocessing steps (missing data handling, scaling or not, encoding of categorical variables), train/test split or cross-validation strategy, hyperparameter tuning procedure (e.g., grid search ranges, final selected values), and any steps taken to avoid overfitting.

While the authors mention the potential of enhancing self-efficacy and health literacy through interventions, the implications for practice and policy remain somewhat generic; it would strengthen the manuscript to propose more concrete, context-sensitive recommendations, such as specific nurse- or physician-led programs, integration with digital/e-health literacy training in outpatient clinics, considerations of health-system constraints in Indonesia (time, staffing, technology), and suggestions for how clinicians might use model-derived risk profiles to triage or tailor support.

Authors revised the manuscript and uploaded the document.

## 1.2. Reviewer 2

Reviewer:

The outcome definition and modeling approach require clearer justification: it is not fully explicit whether adherence was treated as a continuous variable or categorized into levels, and why this operationalization was chosen; the authors should state the exact form of the dependent variable used in the Random Forest, report appropriate performance metrics for that type of outcome (e.g.,  $R^2$  and RMSE for regression or AUC, sensitivity, specificity, and calibration indexes for classification), and discuss whether these performance indices are clinically meaningful rather than only statistically adequate.

The Results section presents intuitive and theoretically coherent correlations between adherence, self-efficacy, and health literacy, but the interpretation currently leans toward causal language in several places; given the cross-sectional design, the authors should consistently rephrase such statements to reflect associative and predictive relationships only, avoid implying temporal precedence, and explicitly acknowledge that the directionality between psychological constructs and adherence cannot be inferred from these data.

The feature importance outputs from the Random Forest are one of the most interesting contributions of the study, yet they are somewhat underdeveloped in both reporting and interpretation; I recommend presenting a ranked list or plot of variable importance with confidence intervals or stability checks (e.g., across resamples), distinguishing health literacy subdomains, self-efficacy facets, and sociodemographic factors, and then more deeply discussing how these specific patterns can inform targeted intervention design for women with chronic illness in Indonesian public hospitals.

The Discussion appropriately highlights the clinical relevance of moderate levels of adherence, health literacy, and self-efficacy, but it would benefit from a more critical and nuanced appraisal of the model's limitations and potential biases—for example, the exclusive reliance on self-report measures (including adherence), possible social desirability effects, non-coverage of women with severe cognitive/psychiatric problems, and the contextual specificity of urban tertiary-care settings—so that readers have a realistic sense of how and where these findings might translate into practice.

Authors revised the manuscript and uploaded the document.

## 2. Revised

Editor's decision: Accepted.

Editor in Chief's decision: Accepted.