

Predicting Fertility Intentions Among Young Adults Using Machine Learning: The Relative Importance of Family Expectations, Housing Insecurity, Gender Ideology, Career Uncertainty, Relationship Quality, and Cultural Values

Katarína. Horváthová¹, James. O'Connor^{2*}

¹ Department of Psychology, Comenius University in Bratislava, Bratislava, Slovakia

² Department of Health Psychology, University of Manitoba, Winnipeg, Canada

* Corresponding author email address: james.oconnor@umanitoba.ca

E d i t o r	R e v i e w e r s
Zahra Yousefi  Assistant Professor, Department of Psychology, Isfahan Branch (Khorasgan), Islamic Azad University, Isfahan, Iran Z.yousefi1393@khuif.ac.ir	Reviewer 1: Mehdi Ghezelsfloo  Assistant Professor, Department of Psychology, Gonbadkavos University, Gonbadkavos, Iran. Email: m.ghezelsfloo@gonbad.ac.ir Reviewer 2: Mansour Abdi  Associate Professor, Department of Psychology, Arak University, Iran. m.abdi@arak.ac.ir

1. Round 1

1.1. Reviewer 1

Reviewer:

The Introduction emphasizes family expectations, housing insecurity, gender ideology, career uncertainty, relationship quality, and cultural values, but the theoretical integration of these domains remains underdeveloped. The paragraph beginning “Although previous research has identified numerous correlates of fertility intentions” justifies machine learning methodologically but does not provide a unified substantive framework. The authors should organize the predictors using an established fertility decision-making theory, such as the Theory of Planned Behavior, life-course theory, the gender-equity framework, or a multilevel model of fertility intentions. Each domain should be explicitly mapped onto attitudes, subjective norms, perceived behavioral control, relational resources, or structural constraints. This would prevent the model from appearing as a data-driven aggregation of variables without a coherent causal or theoretical architecture.

In the Study Design and Participants section, the manuscript reports that “a total of 1,463 individuals accessed the online survey” and that the final analytical sample included exactly 1,284 participants. The recruitment process requires a more

transparent participant-flow description. The authors should explain whether individuals who accessed the survey but did not provide consent were included in the initial count, whether exclusion rules were established before data inspection, and whether the 92 incomplete questionnaires overlapped with the 31 ineligible respondents or other exclusion categories. A flow diagram or exact sequential accounting should be provided so that the sum of exclusions can be verified and selection bias can be assessed.

The cultural-values measure requires substantial conceptual refinement. The manuscript states that the 16-item instrument combines “familism, intergenerational obligation, individual autonomy, religious or moral valuation of parenthood, pronatalist beliefs, acceptance of voluntary childlessness, preference for self-realization, and the perceived social importance of marriage and family continuity.” These dimensions may not form a single coherent continuum, and reverse-coding autonomy or acceptance of childlessness could incorrectly imply that these attitudes are exact opposites of family centrality. The authors should test a multidimensional factor model and avoid reporting a single total score unless unidimensionality is clearly supported. Domain-level importance should preferably be calculated separately for distinct cultural-value subscales.

The handling of relationship quality for unpartnered participants needs further methodological scrutiny. The manuscript reports that structurally missing relationship-quality scores were imputed and accompanied by a binary relationship-status indicator. Even with a missingness indicator, median imputation may create an artificial reference group and allow the model to treat unpartnered respondents as having average relationship quality. A more defensible approach would involve stratified models for partnered and unpartnered participants, interaction terms between relationship status and relationship-quality variables, or algorithms that explicitly handle structural missingness. The authors should demonstrate through sensitivity analyses that the ranking of relationship quality is not an artifact of coding decisions.

Response: Revised and uploaded the new document.

1.2. Reviewer 2

Reviewer:

The inclusion criterion stating that participants had to indicate that “having a biological, adopted, foster, or socially parented child remained a personally meaningful or conceivable future option” creates a potentially serious selection problem. By excluding respondents who definitively rejected all pathways to parenthood, the study may systematically remove individuals with the strongest negative fertility intentions. This restriction could distort the outcome distribution, reduce external validity, and artificially improve classification performance. The authors should justify this criterion theoretically, report how many respondents were excluded for this reason, and preferably repeat the analysis in a sample that includes voluntarily childfree individuals. Fertility-intention research should generally retain definite non-intenders unless the outcome is explicitly restricted to timing among those open to parenthood.

In the Participants subsection, the manuscript states that “stratified recruitment procedures were used to obtain participants from all 16 German federal states.” More information is needed about the sampling frame, panel provider, recruitment quotas, and achieved representativeness. The authors should compare the sample with official German population distributions for age, gender, education, migration background, employment, and federal state. The present sample appears highly educated, with 67.3% holding or pursuing tertiary qualifications, which may limit generalizability. If post-stratification weights were not used, the authors should explain why and acknowledge that the predictive model reflects the recruited sample rather than the national population of German young adults.

The manuscript reports an a priori minimum sample requirement of 1,100 cases but provides no reproducible basis for this number. The statement that the sample was sufficient for “comparing multiple classification algorithms, tuning model hyperparameters, estimating out-of-sample performance, and calculating model-agnostic variable-importance indices” is too general. The authors should report the expected prevalence of positive fertility intention, the anticipated number of candidate predictors after dummy coding, the planned training-to-test split, and the minimum number of outcome events per validation fold. A learning-curve analysis would also help demonstrate whether model performance stabilized at the available sample size.

In the Data Collection Tools section, the fertility-intention outcome was created from three items assessing desire, intention to take concrete steps, and perceived likelihood. These indicators represent conceptually distinct constructs, and averaging them may obscure meaningful discrepancies between wanting a child, planning for one, and expecting one. The authors should provide the exact item wording, response anchors, factor loadings, inter-item correlations, and evidence that a unidimensional score is psychometrically defensible. A confirmatory factor analysis should be reported specifically for the outcome measure, and sensitivity analyses should examine whether the machine-learning results differ when desire, intention, and perceived likelihood are modeled separately.

The sentence “participants were categorized as having a positive fertility intention when their mean score was 3.50 or higher” requires stronger justification. The threshold appears researcher-defined and may materially determine class prevalence, model performance, sensitivity, specificity, and importance rankings. The authors should explain whether 3.50 corresponds to an established validation criterion, a scale midpoint, or a clinically or demographically meaningful distinction. The threshold must not be chosen after examining outcome distributions or predictive results. The manuscript should report the complete distribution of fertility-intention scores and present sensitivity analyses using alternative thresholds, ordinal classification, and continuous prediction.

Several predictor scales appear to have been constructed specifically for this study, including the measures of family expectations, housing insecurity, career uncertainty, and cultural values. The statement that “content validity of the questionnaire was evaluated by a multidisciplinary panel” is insufficient for establishing measurement validity. The authors should provide the item-generation procedure, theoretical sources, expert-rating indices, pilot revisions, factor structure, standardized factor loadings, average variance extracted, composite reliability, and discriminant-validity evidence. Because machine-learning models can exploit measurement artifacts, the psychometric quality of each feature set is essential. The complete instrument should be included as supplementary material in German and, where relevant, in an English translation.

Response: Revised and uploaded the new document.

2. Revised

Editor’s decision: Accepted.

Editor in Chief’s decision: Accepted.