Power Over Presence: Replication and Model Evaluation

Nina Rice & Erika Garza-Elorduy

2025-03-15

Power Over Presence: Women's Representation in Comprehensive Peace Negotiations and Gender Provision Outcomes

A Replication Project by Nina Rice and Erika Garza-Elorduy

Abstract

This study replicates and extends Elizabeth Good's analysis on the impact of women's roles in peace negotiations on gender provisions in agreements. Using data from 116 peace agreements and 2,299 Track 1 delegates, we replicate her OLS models with corrected standard errors and assess robustness using Firth logistic regression to address rare-event bias. We also evaluate the predictive performance of OLS and logistic models using cross-validation and extend the analysis by incorporating time and country-level effects. Our findings confirm that only women in high-power roles, such as signatories, significantly influence gender provisions. This replication underscores the importance of power over presence while highlighting methodological refinements that improve model reliability.

1. Introduction

Elizabeth Good's study, *Power Over Presence: Women's Representation in Comprehensive Peace Negotiations and Gender Provision Outcomes*, examines how the presence of women in peace negotiations affects the inclusion of gender provisions in peace agreements. Using data from 116 comprehensive peace agreements finalized between 1990 and 2021, along with information on 2,299 Track 1 delegates, the study categorizes women's roles as signatories, negotiators, mediators, or observers. Good's findings challenge the assumption that simply increasing the number of women in negotiations leads to stronger gender outcomes. Instead,

her analysis suggests that only women in positions of authority, such as signatories, significantly influence the inclusion of gender provisions. The study controls for factors such as conflict-related sexual violence, international involvement, and national gender equality, reinforcing the argument that power matters more than mere presence in shaping gender-responsive peace agreements.

This replication project seeks to verify Good's findings by reproducing her core analyses and assessing the robustness of her conclusions. Our approach follows her original methodology, including the use of OLS regression models to examine the relationship between women's participation and the inclusion of gender provisions. Additionally, we introduce alternative modeling techniques, such as logistic regression and Firth regression, to evaluate whether different statistical approaches yield consistent results. Through these models, we aim to determine whether the effects Good identifies remain stable across various analytical frameworks.

Beyond replication, we extend the original analysis by incorporating time-based and country-based factors to explore how the inclusion of gender provisions in peace agreements varies across different contexts. We use cross-validation techniques to compare the predictive performance of the models and assess their generalizability. By testing the assumptions and methodological choices of the original study, we aim to refine our understanding of the relationship between women's roles in peace negotiations and the outcomes of gender provisions.

This project contributes to the broader discussion on women's substantive representation in peace processes by examining whether positions of influence, rather than mere presence, shape policy outcomes. Through replication and model evaluation, we assess the strength of Good's conclusions and provide insights into potential limitations and areas for further research.

2. Summary of the Original Paper

2.1. Unit of Analysis and Sampling

The unit of analysis in this study is comprehensive peace agreements, specifically the 116 agreements finalized between 1990 and 2021. Each agreement serves as a single case, meaning the analysis is conducted at the agreement level rather than at the individual delegate level. However, because the study examines how women's involvement influences the inclusion of gender provisions, the dataset also includes information on the composition of delegates within each agreement as independent variables. The Peace Agreement Delegate Dataset (PADD) provides data on 2,299 Track 1 delegates, categorizing them by role, such as signatory, negotiator, mediator, or observer, and linking them to their respective agreements. The study assesses whether agreements with women in key roles are more likely to include gender provisions, using agreement-level data to identify broader patterns.

There are valid concerns about observational independence due to the nature of peace agreements. Many agreements in the dataset are part of broader peace processes, meaning that successive agreements may be influenced by earlier negotiations within the same conflict. This

creates potential dependencies between observations, as agreements from the same conflict may share similar characteristics, actors, or external pressures. Additionally, individual delegates may participate in multiple agreements, leading to non-independence across cases where the same key negotiators shape different agreements. This clustering effect could result in correlated observations, requiring statistical adjustments such as robust standard errors or multilevel modeling to account for within-conflict dependencies.

The data were obtained from a comprehensive set of sources rather than through random sampling. The study includes all available comprehensive peace agreements from 1990 to 2021, ensuring that the dataset represents the full population of relevant cases rather than a subset. Agreements were identified and coded using sources such as the PA-X Peace Agreement Database, the UN Peacemaker Database, and direct analysis of agreement texts. The study codes all Track 1 delegates mentioned in these agreements, using name-based gender identification techniques supplemented by manual verification to ensure accuracy. This process involves drawing on databases and algorithms that associate names with gender, as well as cross-checking biographical information from official records, media sources, and expert knowledge to reduce misclassification errors. While this approach allows for a comprehensive assessment of women's participation in peace processes, it also means that the dataset reflects the documentation and availability of agreements rather than a probabilistic sample.

2.2. Dependent Variable (`GeWom`)

The dependent variable for the model being replicated is GeWom, which indicates whether a comprehensive peace agreement includes provisions specifically addressing women. It is a binary variable coded as:

- 0 = The agreement does not contain provisions for women.
- 1 = The agreement does contain provisions for women.

This variable serves as a measure of women's substantive representation in peace agreements by identifying explicit references to women, girls, widows, mothers, gender-based violence, or sexual violence within the agreement text. By coding the presence or absence of such provisions, the variable captures the extent to which women's issues are formally included in peace processes.

The dataset originally consists of 116 observations, corresponding to 116 comprehensive peace agreements finalized between 1990 and 2021. However, due to missing delegate data in some cases, the final analytical sample is reduced to 108 agreements. This adjustment ensures that only agreements with complete delegate information are included in the analysis, maintaining data integrity while slightly reducing the overall sample size.

The distribution of GeWom, visualized in the bar chart, reveals that a majority of agreements include gender provisions, with nearly 58.6% coded as 1 and around 41.4% coded as 0. This

suggests that while gender considerations are present in most agreements, a significant proportion still lack explicit provisions for women. The near-even split between the two categories ensures a reasonable balance for model estimation, reducing concerns about class imbalance that could affect predictive performance.

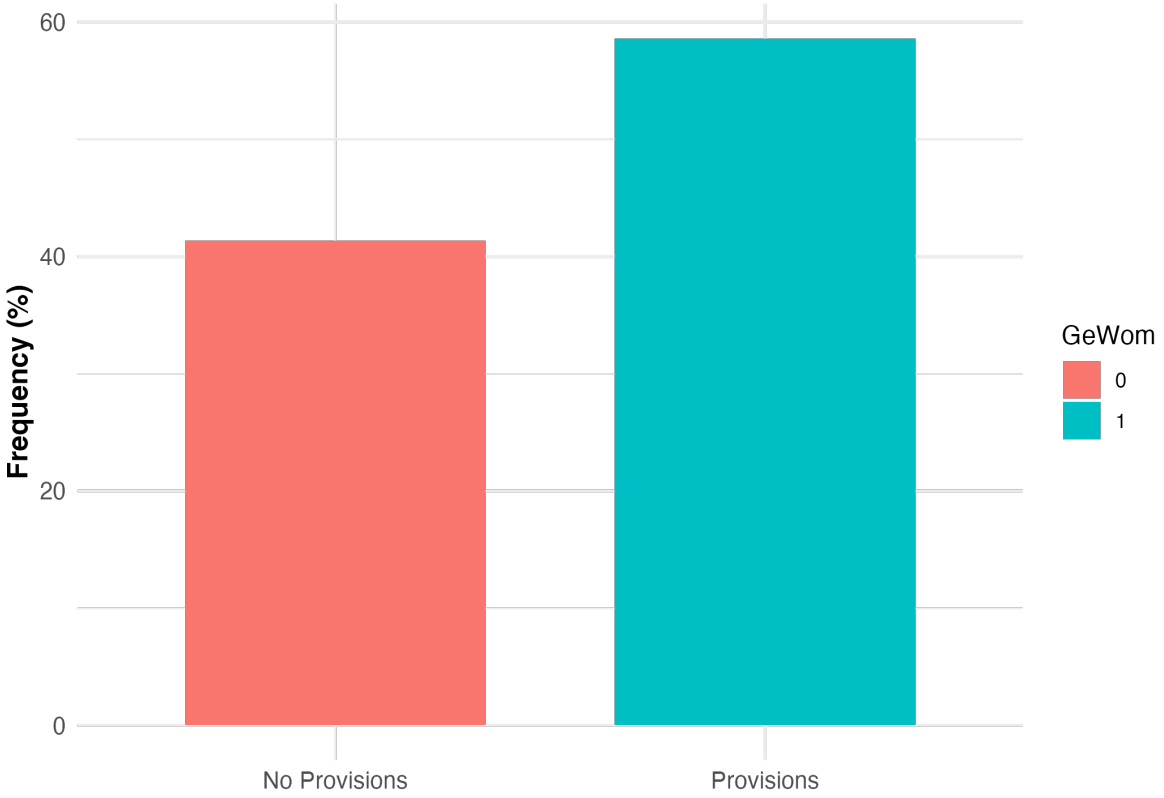

Figure 1: Proportion of peace agreements with (GeWom = 1) and without (GeWom = 0) gender provisions.

2.3. Missing Data & Handling Strategy

The authors addressed missing data using multiple imputation with the mice package in R, creating five imputed datasets for their main analyses. However, for the four OLS models in Table 4, they did not use imputed data. Instead, they employed casewise deletion, meaning that any agreements with missing values in the relevant variables were excluded from the analysis. This decision results in a smaller but complete dataset for the OLS models, ensuring that missing values do not affect the model estimates.

The missingness map provided visually represents the extent and pattern of missing data

across variables. The dataset has 15.5% missing values, as indicated by the proportion of black areas in the visualization, with some variables exhibiting more missing data than others. This missingness could introduce bias if the missing values are not randomly distributed, reinforcing the importance of either imputation or careful casewise deletion.

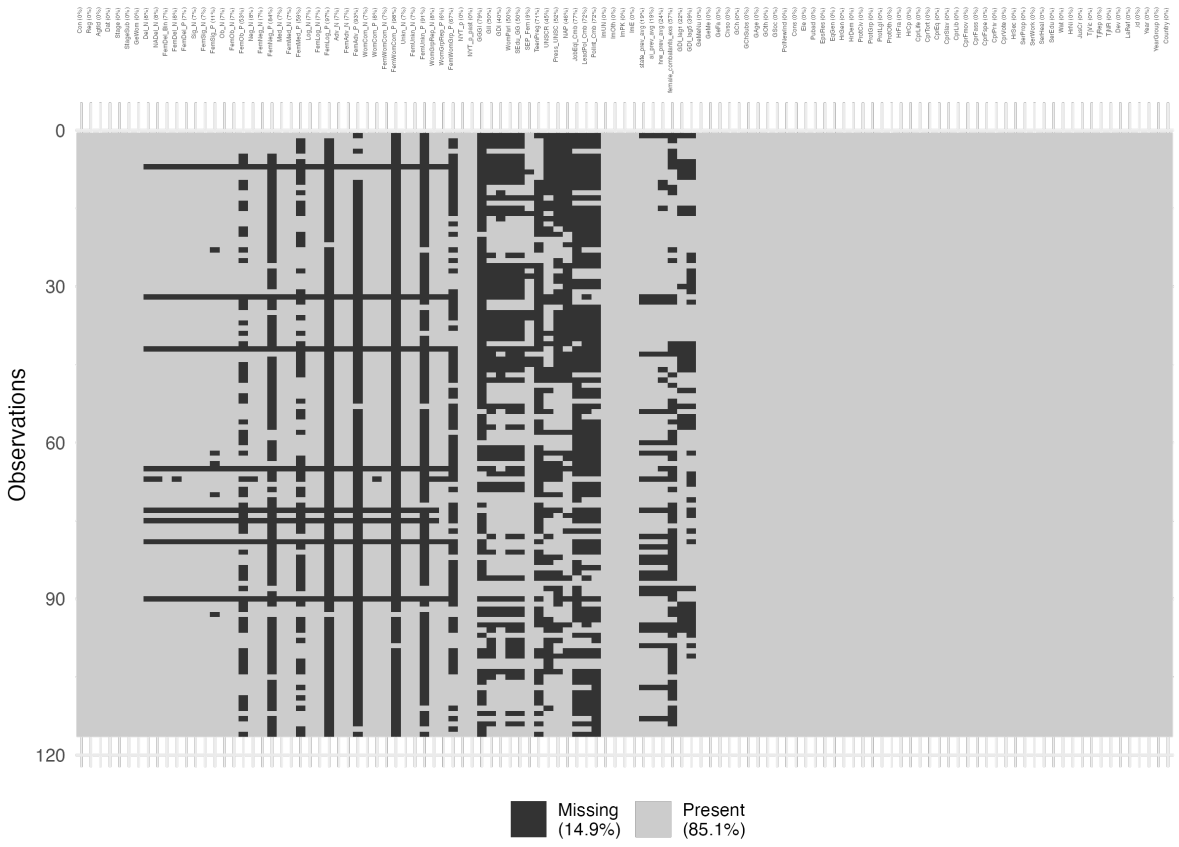

Figure 2: Missing data visualization showing the proportion and distribution of missing values across variables in the dataset.

2.4. Original Model

The models we are replicating are Ordinary Least Squares (OLS) regressions, specifically four separate models from Table 4 of the original paper. These models analyze the relationship between different categories of women delegates—signatories, negotiators, mediators, and observers—and the inclusion of gender provisions in comprehensive peace agreements. The regression estimates the linear relationship between the percentage of women in each role and the likelihood of gender-related content being included in the agreements. Additionally, we applied robust standard errors (HC3) to correct for potential heteroskedasticity.

The original author aimed to establish causal inference rather than simple description or prediction. Their goal was to assess whether women's presence in peace negotiations influences substantive outcomes, particularly the inclusion of gender provisions. They challenge the assumption that simply increasing women's participation leads to improved representation, arguing instead that only women in high-power roles (e.g., signatories) significantly impact negotiation outcomes.

3. Replication Results

We successfully replicated the four OLS models from Table 4. Our coefficient estimates, standard errors, and the number of observations match those in the original paper, confirming that our replication accurately follows the original methodology and dataset. Furthermore, we applied corrections for standard errors using robust variance estimation to account for heteroscedasticity, confirming that the relationship between women's roles and gender provisions remains statistically significant in key cases.

Percentage Measurement of Women Delegates (OLS Regression)

	Dependent variable:			
	Provisions for Women (GeWom)			
	(1)	(2)	(3)	(4)
Women Signatories	2.096**			
	(0.898)			
Women Negotiators		2.352**		
		(0.899)		
Women Mediators			0.259	
			(0.799)	
Women Observers				0.496
				(0.508)
Constant	0.517***	0.277**	0.639***	0.595***
	(0.052)	(0.122)	(0.073)	(0.070)
Residual Std. Error	0.488	0.441	0.488	0.491
F-statistic	5.444	6.845	0.105	0.953
Observations	103	18	48	55
R2	0.051	0.300	0.002	0.018
Adjusted R2	0.042	0.256	-0.019	-0.001
Residual Std. Error	0.488 (df = 101)	0.441 (df = 16)	0.488 (df = 46)	0.491 (df = 53)
F Statistic	5.444** (df = 1; 101)	6.845** (df = 1; 16)	0.105 (df = 1; 46)	0.953 (df = 1; 53)
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure 3: Replicated OLS regression results confirm original findings regarding the impact of women's negotiation roles on gender provisions in peace agreements.

When comparing coefficients and goodness-of-fit metrics, all four OLS models did not perform strongly. The OLS regression for women signatories (FemSig_P) had an R-squared value of 0.103, indicating that only 10.3% of the variation in GeWom is explained by the model. The root mean squared error (RMSE) was 0.4897, suggesting a high level of error in the model's predictions, and the mean absolute error (MAE) was 0.476. The OLS regression for women negotiators (FemNeg_P) had an even lower R-squared value of 0.079, suggesting weak explanatory power, and similar error metrics. The OLS model for women mediators (FemMed_P) yielded an R-squared of 0.0002, indicating almost no relationship, while the model for women observers (FemOb_P) had an R-squared of 0.011, suggesting a negligible effect. Across all models, the low R-squared values and high prediction errors indicate that OLS is not well-suited for predicting a binary outcome like GeWom.

4. Alternative Models

4.1 Model Justification

The additional models used in this analysis include Firth logistic regression, a logistic regression model with a time-based variable, and a logistic regression model with a country-based variable. The Firth logistic regression was applied to the same dependent variable, GeWom, and was run separately for four independent variables: the percentage of women signatories (FemSig_P), negotiators (FemNeg_P), mediators (FemMed_P), and observers (FemOb_P). In addition, a logistic regression model was used to assess the impact of FemSig_P over different time periods using YearGroup as a categorical variable, and another logistic regression model was run to examine the relationship between FemSig_P and GeWom while accounting for country-level differences by including Country as a factor.

These models were chosen for different methodological reasons. Firth logistic regression was used because the dataset is small, and standard logistic regression can suffer from separation issues, leading to inflated estimates. Firth regression applies a penalized likelihood approach that improves coefficient stability. The time-based logistic regression model was used to test whether the relationship between FemSig_P and GeWom changed over different decades, as historical trends may influence the inclusion of gender provisions. The country-based logistic regression model was used to assess whether the effects of FemSig_P on gender provisions vary by country, as local political and cultural factors may shape negotiation outcomes.

4.2 Firth regressions GeWom ~ FemSig_P, GeWom ~ FemNeg_P, GeWom ~ FemMed_P, GeWom ~ FemOb_P

The results from the four Firth regression models show that the coefficient for women signatories (FemSig_P) is 10.22 with a standard error of 5.13, a 95% confidence interval of [1.40, 22.91], and a p-value of 0.0202. This suggests a statistically significant positive relationship between FemSig_P and GeWom at the 5% level. Women negotiators (FemNeg_P) also have a strong positive coefficient of 10.51, but with a larger standard error of 6.00, a 95% confidence interval of [1.03, 35.00], and a p-value of 0.0261. Although this result is statistically significant at the 5% level, the wide confidence interval indicates a high degree of uncertainty. Women mediators (FemMed_P) have a coefficient of 0.49 with a standard error of 3.06, a 95% confidence interval of [-5.31, 7.88], and a p-value of 0.8714, which is not statistically significant. Similarly, women observers (FemOb_P) have a coefficient of 1.65 with a standard error of 2.23, a 95% confidence interval of [-2.19, 8.60], and a p-value of 0.4321, indicating a lack of statistical significance. The likelihood ratio test results align with these findings, confirming that only women signatories and negotiators have a meaningful association with gender provisions in peace agreements.

	Firth's Regression on GeWom			
	(1)	(2)	(3)	(4)
Intercept	0.055	-0.912	0.560 *	0.382
	(0.214)	(0.589)	(0.309)	(0.290)
Women Signatories	10.219 **			
	(5.126)			
Women Negotiators		10.505 **		
		(6.000)		
Women Mediators			0.492	
			(3.057)	
Women Observers				1.649
				(2.233)

*** p < 0.01; ** p < 0.05; * p < 0.1

Figure 4: Firth logistic regression results address rare-event bias, reinforcing that only women signatories significantly influence the inclusion of gender provisions in peace agreements.

4.3 Logistic regressions GeWom ~ FemSig_P + Date, GeWom ~ FemSig_P + Con

The two logistic regression models we implemented examine the impact of different contextual factors—time and country—on the likelihood that peace agreements include provisions for women while controlling for the presence of women signatories. The first model analyzes the relationship between women signatories and the inclusion of gender provisions over time by grouping agreements into 10-year intervals. The second model investigates whether certain countries are more or less likely to include gender provisions when accounting for the percentage of women signatories.

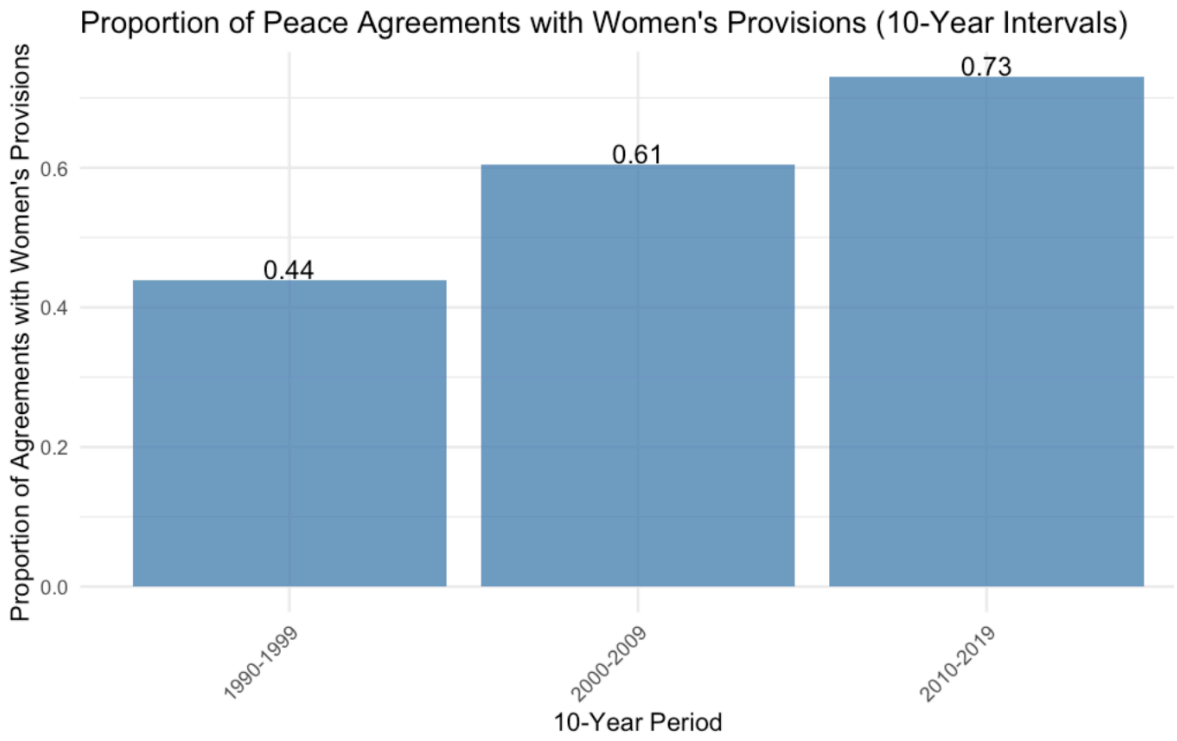

Figure 5: Proportion of peace agreements including provisions for women over three consecutive decades (1990-2019).

The time-based model shows that a higher percentage of women signatories increases the likelihood of gender provisions being included, with a coefficient of 11.00 and a standard error of 5.83. While this effect is positive and relatively strong, it is only weakly significant with a p-value of 0.081. Additionally, the results indicate that agreements signed between 2010-2019 are significantly more likely to contain gender provisions compared to those from 1990-1999, with a coefficient of 1.01 and a p-value below 0.1. This suggests that the inclusion of gender provisions has increased over time, independent of the percentage of women signatories.

Logistic Regression: Impact of Time on Women's Provisions

	Dependent variable:
	Provisions for Women (GeWom)
Women Signatories	11.004*
	(5.832)
Year Group	0.621
	(0.497)
YearGroup2010-2019	1.013*
	(0.517)
Constant	-0.450
	(0.350)
Observations	103
Log Likelihood	-65.397
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 6: Logistic regression results for the impact of time (YearGroup) on the inclusion of gender provisions in peace agreements.

The country-based model also finds a strong effect for women signatories, with a larger coefficient of 15.33 and a lower standard error, making the effect statistically significant at the $p < 0.05$ level. This suggests that when controlling for country effects, the presence of women signatories remains a key factor in determining whether agreements include gender provisions. However, the model also highlights significant regional disparities. The results indicate that agreements from Colombia are significantly less likely to include gender provisions, with a coefficient of -2.47 and a p-value below 0.1. Conversely, agreements from Sudan and South Sudan display extremely high coefficients, likely due to separation issues in logistic regression caused by a small number of agreements in these countries.

Logistic Regression: Impact of Country on Women's Provisions

	Dependent variable:
	Provisions for Women (GeWom)
Women Signatories	15.333**
	(6.963)
Country	-1.338
	(1.758)
CountryColombia	-2.469*
	(1.442)
CountryDemocratic Republic of Congo	-0.743
	(1.466)
CountryEthiopia	-1.792
	(1.683)
CountryOther	-1.553
	(1.190)
CountrySomalia	-0.115
	(1.637)
CountrySouth Sudan	17.228
	(1,927.224)
CountrySudan	17.467
	(2,917.013)
Constant	1.099
	(1.155)
Observations	103
Log Likelihood	-53.796
Note:	*p<0.1; **p<0.05; ***p<0.01

Comparing

the two models, the country-based model provides a better overall fit, with a log-likelihood value of -53.80 compared to -65.40 in the time-based model. This suggests that country-level variation explains more of the differences in gender provisions than time-based trends alone. However, the country-based model also introduces interpretability challenges, particularly due to extreme coefficients in countries with few agreements. In contrast, the time-based model supports the broader claim that gender provisions in peace agreements have become more common in recent decades.

While these logistic models provide a more appropriate methodological approach than the OLS regressions, they still reveal weaknesses in predictive power. The country-based model improves fit but suffers from instability in cases where observations are sparse. The time-based model is easier to interpret and supports a trend of increasing gender provisions, though its explanatory power is weaker. Future research could explore interactions between time and country effects to assess whether specific regions have driven historical trends in gender provisions.

5. Model Comparison Using Cross-Validation

The original OLS model (GeWom ~ FemSig_P) was evaluated using 10-fold cross-validation to assess its predictive performance. The dataset was split into ten subsets, where the model was trained on nine folds and tested on the remaining fold in each iteration. The cross-validation results produced three key performance metrics: Root Mean Squared Error (RMSE), R-Squared (R^2), and Mean Absolute Error (MAE). The RMSE was 0.4884, indicating that predictions deviated from actual GeWom values by an average of 0.488 units, suggesting poor model performance. The R^2 value was 0.1293, meaning that FemSig_P explained only 12.9% of the variance in the dependent variable, which is slightly higher than the original model's R^2 of 0.051 but still relatively weak. The MAE was 0.4725, meaning that the average absolute difference between predicted and actual values was approximately 0.47 units.

Despite minor improvements in generalizability, OLS regression is not well-suited for binary outcomes. The statistical software flagged a potential issue, suggesting that a classification method such as logistic regression would be more appropriate. Since GeWom is a binary variable, the OLS model likely misrepresents the relationship, as it assumes a continuous outcome. This warning does not invalidate the results but highlights a methodological limitation that suggests other models might be better suited.

Given the issues with using OLS for a binary dependent variable, we also tested alternative models, including logistic regression and Firth regression, which are more appropriate for classification tasks. The logistic regression model, which included time-based (GeWom ~ FemSig_P + YearGroup) and country-based (GeWom ~ FemSig_P + Country) variations, provided better predictive power. The country-based logistic model had a higher log-likelihood (-53.796 vs. -65.397 for the time-based model), indicating a better fit. Additionally, the coefficient for FemSig_P in the country-based model (15.333, $p < 0.05$) was stronger and more statistically

significant than in the time-based model (11.004, $p < 0.1$), further supporting its explanatory power.

Among all the models tested, the country-based logistic regression model performed the best in terms of predictive accuracy and statistical significance. The inclusion of country-level effects explained more of the variation in GeWom than time-based trends alone. However, some country coefficients exhibited extreme values due to data sparsity, which may indicate numerical instability. Despite this limitation, logistic regression was more suitable than OLS for analyzing a binary outcome and provided a clearer interpretation of how women signatories impact gender provisions in peace agreements.

For the best-performing model, the independent variable FemSig_P showed a significant positive effect on GeWom. In the country-based logistic regression, an increase in the percentage of women signatories was associated with a substantial increase in the odds of including gender provisions. This finding reinforces the argument that women's representation in high-power roles, rather than mere presence, is critical for influencing substantive policy outcomes in peace agreements. However, the results also highlight that country-level factors play an essential role in shaping these outcomes, with some nations systematically less likely to include gender provisions. Future research should explore potential interactions between time and country effects to further refine these conclusions.

6. Conclusion

Our replication and extension of *Power Over Presence: Women's Representation in Comprehensive Peace Negotiations and Gender Provision Outcomes* largely support the original study's key findings while identifying some methodological considerations that affect the robustness of its conclusions. The original paper argues that women's mere presence in peace negotiations does not necessarily lead to gender provisions in agreements. Instead, only women in high-power roles, such as signatories, significantly increase the likelihood of these provisions being included. Our replication of the OLS models confirmed this general trend, but we found that OLS is not the most appropriate approach for modeling a binary outcome like GeWom. Logistic and Firth regressions provided more appropriate analytical frameworks, reinforcing the finding that women signatories play the most substantial role in influencing peace agreements. However, we also found that country-level differences play an essential role, with some nations systematically less likely to include gender provisions.

We are moderately confident in the authors' conclusions, as our analysis supports the claim that power matters more than presence in gender-responsive peace negotiations. However, our findings suggest that the study could benefit from refinements in its modeling approach. The use of OLS regression for a binary dependent variable introduces interpretational challenges, and while the logistic and Firth models improved predictive accuracy, the overall explanatory power remained limited. Additionally, our cross-validation results indicate that the percentage of women signatories, while significant, is not a particularly strong standalone predictor of

gender provisions. This suggests that other unaccounted-for variables may play a role in shaping these outcomes.

A key area for further exploration would be to expand the analysis beyond individual agreements to consider broader political and structural factors that influence the inclusion of gender provisions. For example, future research could examine whether the presence of women in leadership positions within negotiating parties, rather than just among the signatories, contributes to gender-sensitive outcomes. Additionally, the study could be strengthened by incorporating interaction terms to explore whether the impact of women signatories varies depending on contextual factors such as international involvement or conflict intensity.

The primary limitations of this study include the relatively small sample size, the potential non-independence of observations due to repeated agreements within the same conflict, and the high variability in the data, which affects the stability of some model estimates. Additionally, our results highlighted potential issues with separation in logistic regression models, particularly when including country-level effects. To address these challenges, future work could explore alternative methods such as hierarchical modeling or Bayesian approaches to better account for dependencies in the data.

Overall, while our replication confirms the core argument that women in positions of power have a meaningful impact on gender provisions in peace agreements, our findings also suggest that the study's conclusions should be interpreted with caution due to methodological constraints. Moving forward, refining the modeling approach and incorporating additional contextual factors would help strengthen the evidence base for understanding the role of women in peace negotiations.

7. References

Good, Elizabeth. 2024. "Women's Roles in Peace Negotiations and Gender Provisions." *International Studies Quarterly*68(2): 215–30.

AI Disclosure Statement

We used ChatGPT 4o in this project to help with R code and statistical interpretation. Doing so saved us time with troubleshooting errors, syntactical assistance, and pulling our various model results together. We found this to be beneficial for completing the project on a brief timeline.