The limits of gender quota in Indonesia: A quasi-experimental approach

Noory Okthariza

Data

I leverage provincial-level data originating from the Indonesian Election Commission (KPU). Before the campaign season begins, all candidates are required to submit resumes containing information about their party affiliation, list position, constituency, gender, province, and occupation – among other details. I scraped these resumes from the KPU website, collecting 32,877 pre-cleaned observations. Additionally, I enriched this dataset with information on district magnitude, also obtained from the KPU website. Data on elected candidates was derived from the official KPU recap, which was made available approximately one month after the election.

There were 18 parties competed nationally in the 2024 election. There were additions of six local parties in the province of Aceh, which is part of Indonesia's special regional autonomy scheme. I excluded these local parties from my analysis to maintain comparability. The list position variable indicates where candidates are placed on the ballot paper. Indonesia utilizes an open-list proportional representation system where the candidate's name, photo, party, and list position are visible to voters on the ballot. The list position can vary from 1 to 12, depending on the size of the constituency and the party's perceived competitiveness. A higher list position—closer to zero—suggests different competitive advantages. Typically, candidate with a smaller number or higher list is associated with incumbency status and various voter heuristics, which can substantially enhance a candidate's competitiveness (Jankowski & Müller, 2021). Additionally, the district magnitude, which depends on the population size, ranges from 3 to 12 seats.

The variable on gender quota compliance is calculated by dividing the number of women nominated by each party in a constituency by the total number of candidates. Each party is expected to meet a 30 percent quota requirement. However, enforcement of this policy is weak, and the KPU has issued a new regulation that allows for rounding down calculations when the result is less than a 0.5 decimal point. Consequently, we can expect parties to sometimes ignore the quota rule, particularly in constituencies with 4, 5, 7, and 8 seats. For instance, a party may choose to nominate only one woman in a constituency with five seats (0.3 x 5 = 1.5) or two women in a district with eight seats (0.3 x 8 = 2.4), instead of rounding up to the nearest whole number. This potential lack of adherence to the quota rule provides a unique opportunity to examine the impact of varying levels of compliance on electoral outcomes. I exploited this variability as a treatment assignment. Accordingly, candidates from parties that nominate at least 30 percent women are categorized as the 'treated' group, while those from parties nominating fewer than 30 percent are considered the 'untreated' group. To ensure covariate balance is achieved in the pre-treatment state, both the mean differences and visual inspections of adjusted and unadjusted samples are provided in the appendix.

Method

While omitted variable bias can be addressed by carefully selecting relevant covariates, one of the pressing empirical challenges in my quasi-experimental design is the presence of potential confounders. It is possible that among the compliers, smaller parties perform better in meeting the gender quota, potentially signaling to voters their stronger commitment to women's representation. Furthermore, the likelihood of women being elected may be enhanced if constituencies comprise both progressive and conservative areas, with voting patterns favoring women more pronounced in progressive constituencies that value minority groups over conservative ones. If this is the case, then our treatment assignment might be biased, as these two possibilities could significantly influence the overall selection of the treatment group.

To address this challenge, I use propensity score matching (PSM) with the nearest neighbor method, which allows me to pair groups with the closest baseline propensity scores to mimic a full experimental design. I also incorporate Mahalanobis distance to refine the matching quality and achieve covariate balance (see Ghorbani, 2019). The PSM design is preferable because the nature of the data does not permit multiple time periods and the parallel trends often assumed in other quasi-experimental designs, such as difference-in-differences and regression discontinuity (Hausman & Rapson, 2018; Caliendo & Kopeinig, 2008). Likewise, it does not require me to defend a challenging assumption about the exclusion restriction commonly found in instrumental variable designs (Angrist et al., 2012). Therefore, once covariate balance is achieved, the validity of the design should rest on the selection and credible measurement of covariates—as previously discussed—and the procedures used to calculate the baseline propensity. I predicted the baseline propensity of receiving treatment using the following logit model:

$$\log\left(\frac{P(Y_i = 1 \mid X_i)}{1 - P(Y_i = 1 \mid X_i)}\right) = \beta_0 + \beta_1 \text{constituency}_i + \beta_2 \text{dm}_i + \beta_3 \text{party}_i + \beta_4 \text{sex}_i + \beta_5 \text{list position}_i \tag{1.1}$$

The explanation of the model rests on the following arguments. First, the probability of receiving treatment ($Y_i = 1$) increases when a specific set of covariates, X_i, are present. For example, in a relatively inclusive constituency (β_1), women tend to be more competitive, which may incentivize party leaders to nominate more women, thereby increasing the likelihood of complying with the quota rule. Second, quota compliance also depends on the size of the district (β_2), as a larger district magnitude provides more seats available for women. Third, party initiative (β_3) should also be accounted for, as parties with a higher commitment to the quota policy might be more inclined to nominate women. Fourth, I argue that gender (β_4) and list position (β_5) are critical components for quota compliance. This is because the decision to nominate women partly depends on the supply of women candidates themselves, as well as the availability of list positions within each party. Parties may nominate women merely for the sake of tokenism, placing them in less competitive positions on the ballot. Alternatively, they may demonstrate a good faith by positioning women in more competitive spots, such as higher on the list (e.g., list numbers 1 or 2).

To evaluate the impact of our treatment variable, I run two different models: one without and the other with covariate controls. It is important to note that the focus of the analysis is on the causal estimation of the average treatment on the treated (ATT), not on the entire sample or the average treatment effect (ATE). Although this approach affects the sample size, it enables a more focused impact evaluation of the quota policy. The full model with covariates is as follows:

$$Y_i = \beta_0 + \beta_1 \cdot \text{treatment}_i + \beta_2 \cdot \text{list position}_i + \beta_3 \cdot \text{sex}_i + \beta_4 \cdot \text{dm}_i$$

(1.2)

To account for the impact of the policy on the entire sample, two additional regressions will be employed. The first is a standard logistic regression conducted without adjusting for the propensity scores of covariates. This estimate is likely to be biased, given that the presence of confounders can significantly affect the error term. But this approach is undertaken to observe how the PSM model differs when compared with the standard model. Secondly, we also employ a multilevel approach. This serves as a robustness check for the analysis, considering that each candidate is essentially nested within the hierarchical structure of entities such as parties and provinces. Analyzing individual variables, such as women candidates, and comparing them against aggregated, higher-level data such as parties, provinces, or constituencies may be prone to ecological fallacy if careful attention is not paid to the established relationships between individual and group-level variables (Gelman & Hill, 2006). This problem arises as we often attempt to draw interpretations for sub-unit level data with a larger N from higher unit data with fewer values or N. As a result, we essentially interpret from less information and different levels of data, which can cause statistical analyses to lose their statistical power due to reduced observations (Greenland, 2000). The multilevel model is run as follows:

$$Y_i = \beta_0 + \beta_1 \text{ treatment} + \beta_2 \text{ list position} + \beta_3 \text{ sex} + \beta_4 \text{ dm} + (\text{party effect} + \epsilon) + (\text{province effect} + \epsilon)$$

(1.3)

The model is similar to 1.2, but we add a random effect for party and province, assuming that candidates from the same party and province tend to be more similar to each other. If that is the case, they may violate the assumption of independence among observations, which is crucial for statistical inference. The errors for party and province are assumed to be normally distributed with a mean of zero and an estimated standard deviation. This is formally executed as follows:

$$\epsilon_{\text{province}} \sim \mathcal{N}(0, \sigma^2) \quad \epsilon_{\text{party}} \sim \mathcal{N}(0, \sigma^2)$$

(1.4)

Lastly, I set two different samples: one with all pooled candidates' sample and the other restricted to female candidates only. This approach helps determine whether the effects of the treatment and control variables are specific to women or part of a broader phenomenon that can be used to interpret general election outcomes.

Descriptive results

Of the 32,877 candidates running in the provincial-level legislative election in 2024, women make up 11,961 (36.3 percent). This percentage is relatively similar to the national-level election, where women constitute around 37 percent of all candidates.

The provincial election comprises 301 districts spread across 38 provinces. If all 18 parties compete in all districts, there should be 5,418 party-district samples. However, some parties may find it challenging to recruit candidates, especially if party elites know that they are not competitive in certain districts. This is particularly common among new parties founded just before the election. As a result, we obtained only 5,006 party-district samples. From this sample, 3,990 party-districts comply with the minimum 30 percent quota nomination, and 1,016 do not. The top three parties with the highest non-compliance rates are Buruh, Gelora, and Nasdem, while the Islamist party PKS performed the best in terms of quota fulfillment, followed by the ruling party PDIP.

Figure 1. Party non-compliance ranking for gender quota

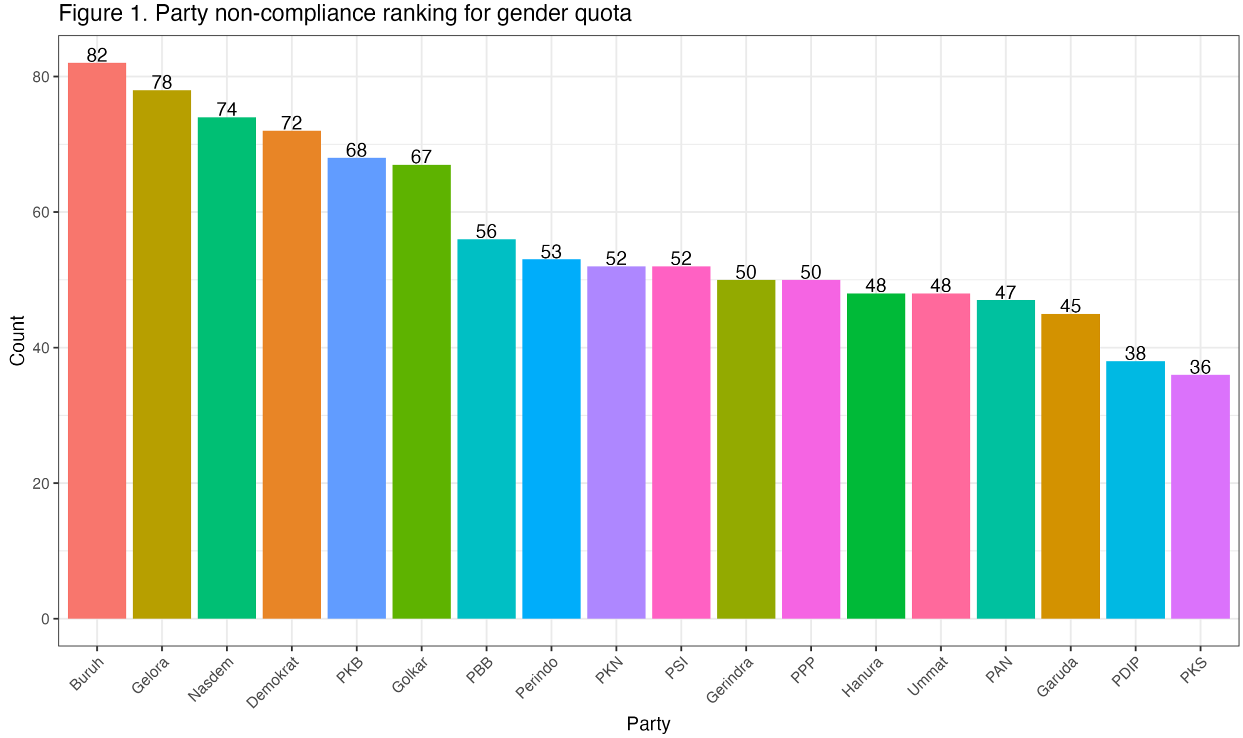
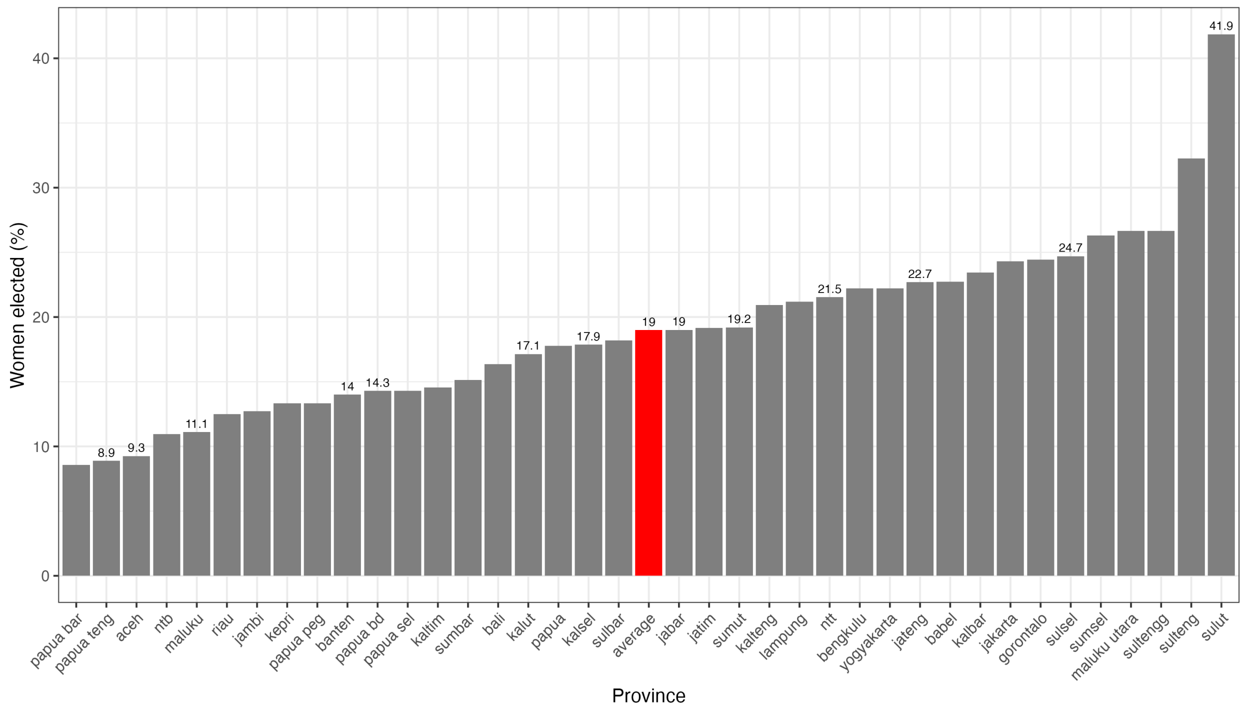


Fig. 2 shows the number of elected women by province. On average, women currently make up 19 percent of legislators across all provinces. This is 3 percent less than in the national parliament (DPR), where women hold 128 out of 580 seats. My findings show that at least 19 provinces are below the average. Two provinces from the eastern part of Indonesia have the lowest (West Papua, 8.6 percent) and highest (North Sulawesi, 41.9 percent) rates.

Education does not seem to matter much in increasing gender representation. Assuming that Javanese provinces are among the most highly educated in Indonesia, the graph suggests that provinces such as West Java, East Java, Yogyakarta, and Central Java are only marginally above average. Another Javanese-based province, Banten, on the other hand, ranks below average, while Jakarta achieves a moderate rate of 24 percent.

Figure 2. Percentage of women elected by province



Continued...

Main results

Despite all the hype surrounding the gender quota campaign, my best assessment is that such a policy has limited, if not unclear, effects on advancing women's political participation in Indonesian provinces. My PSM estimation for all pooled women samples in Table 1 (models 1 and 2) suggests that the null hypothesis is likely true—that is, the treatment does not seem to influence female election outcomes. The standard errors, both with and without covariate controls, are larger than the coefficients, indicating non-significant p-values and high uncertainty around the estimates of the treatment effect. For a simple sensitivity check, I slightly changed the model specification by including a 'gender' variable and incorporating the whole samples for both men and women in models 3 and 4. The results remain consistent, with and without the covariate controls.

As shown in Fig. 3, the likelihood of women being elected slightly improves when the party-district adheres to the 30 percent rule. However, the overall effect of gender compliance is higher for men than for women. Men also tend to have a higher likelihood of getting elected in the untreated group. Regardless, it is evident that the magnitude of the treatment effect remains very small, as the log odds are below 0.05 for both men and women. This suggests that while the quota may influence election outcomes to some extent, its impact is marginal and does not substantially alter the electoral advantages of either gender.

What is clear from the results is that the election largely depends on where candidates are placed on the ballot, district magnitude, and sex. List position shows negative coefficients in all models, suggesting that the higher the list position a candidate secures (i.e., a smaller

Table 1: Propensity score matching results on candidate electability

	Model 1	Model 2	Model 3	Model 4
(Intercept)	−3.114***	−1.870*	−2.522***	−1.601***
	(0.723)	(0.776)	(0.089)	(0.133)
Treatment	−0.117	0.510	−0.035	0.046
	(0.725)	(0.749)	(0.092)	(0.097)
List position		−0.950***		−0.610***
		(0.048)		(0.017)
District magnitude		0.128***		0.125***
		(0.023)		(0.011)
Gender				−0.595***
				(0.060)
Num.Obs.	10 253	10 253	28 195	28 195
AIC	3318.9	2570.7	14 611.7	11 994.5
BIC	3333.4	2599.6	14 628.2	12 035.8
Log.Lik.	−1657.454	−1281.349	−7303.847	−5992.261

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

list number), the more likely they are to be elected. Focusing on model 2, for every one-step move down the list, the odds ratio for a woman candidate decreases by approximately 61.3%, holding all other factors constant (p-value below the 1% level). Conversely, when the same estimation is conducted for the entire sample, the odds of all candidates decrease by 45.7 percent, on average. These significant results confirm prior literature on Indonesia's elections and beyond (Dettman et al., 2017; Lutz, 2010), highlighting the importance of list number in the context of open-list PR elections. However, they also indicate that list position is generally more impactful for women than for men. Therefore, placing women in competitive positions such as list numbers 1 and 2 could significantly improve their chances of winning seats, more so than solely relying on quota policy.

Fig.3. Marginal effect of the treatment based on gender

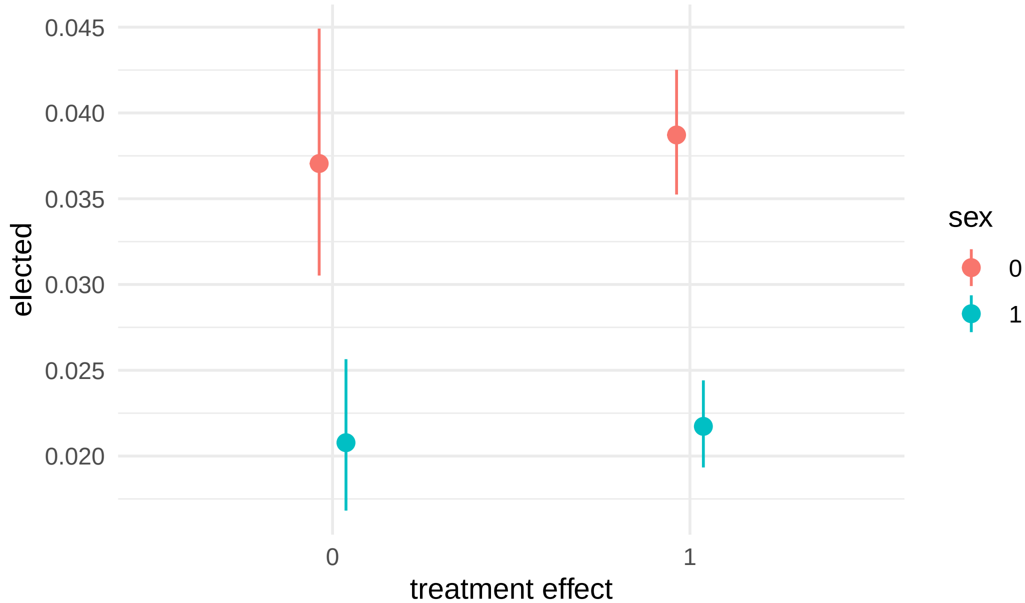
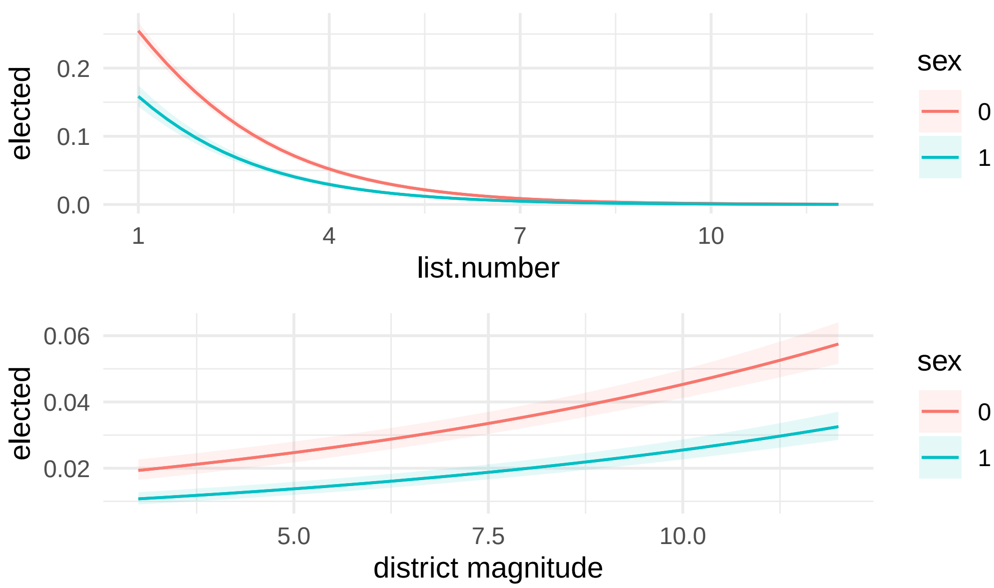


Fig.4. Marginal effects of list position, dm, and gender



I also confirm the widely accepted literature on the impact of larger district sizes (Crisp et al., 2007). Intuitively, the availability of more seats reduces the overall competition, allowing less competitive candidates a better opportunity to secure positions, especially in larger districts compared to more limited-seat scenarios. The binary variable for gender shows negative results, indicating that men, on average, have a higher likelihood of getting elected than women, with a p-value of less than 0.01 percent.

I plot the marginal effects of the covariate controls for list position and district magnitude based on gender in Fig. 4. It shows an inverse relationship for list position, where the probability of success for both men and women exhibit a sudden jump as candidates hold

Table 2: Multilevel model results on candidate electability

	MLM 1 female only	MLM 2 all cand.	MLM 3 all cand.
(Intercept)	−2.923**	−3.935***	−3.177***
	(0.988)	(0.449)	(0.543)
Treatment	0.354	0.152	0.237*
	(0.804)	(0.097)	(0.106)
List position	−1.092***		−0.739***
	(0.053)		(0.018)
District magnitude	0.149***		0.164***
	(0.026)		(0.013)
Gender			−0.613***
			(0.063)
SD (Intercept province)	0.189	0.128	0.186
SD (Intercept party)	2.130	1.819	2.177
Num.Obs.	10 253	28 195	28 195
R2 Marg.	0.477	0.000	0.326
R2 Cond.	0.781	0.503	0.725
AIC	2080.5	13 151.7	9828.0
BIC	2123.9	13 184.7	9885.7

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

a list position closer to zero. This effect remains stronger for men than for women. When candidates are placed in list number five and above, the differential effect diminishes between men and women. On the other hand, the marginal plot for district magnitude shows a positive association, where a higher district magnitude will likely benefit all candidates. The slope for men tends to be steeper, indicating that men are generally more sensitive to larger constituencies than women.

Multilevel model comparison

In this section, I corroborate the previous findings for the treatment using a multilevel model. This method allows us to control for group-level dynamics that might influence our candidate-level data. I argue that the multilevel model is preferable to other possible methods for a sensitivity check because it generally maintains an assumption of linearity between variables of interest (Dedrick et al., 2009), which aligns better with our theoretical expectations and the plotted data in Fig. 4. While other approaches such as quadratic or cubic regression might be suitable for non-linear relationships, this expectation is not supported by the visual inspection of the data and prevailing theories.

Model 1 in Table 2 restricts the analysis to a women-only sample, thereby excluding the binary variable ‘sex’, whereas models 2 and 3 include all candidates. The results in model 1 reveal a null finding for the impact of quota compliance on women’s electoral outcomes. The fixed effect of the treatment variable is not significant, and the standard error is high. However, a random effect variance is likely present. This effect is especially more pronounced between parties than between provinces, as indicated by a higher standard deviation for the party variable. This might suggest that while we observe a null effect of gender compliance on women’s election outcomes, greater variation in quota compliance might be found when disaggregated across different parties, with specific parties potentially exhibiting stronger or weaker compliance effects.

This can further be confirmed in model 2. When we conduct the analysis without covariate controls, the value of R^2 Marginal is zero, possibly indicating a very weak effect of the treatment in explaining the outcome. However, the value of R^2 Conditional is substantially larger, suggesting a significant effect of party and provincial factors in explaining the variability of the outcome. This distinction highlights the importance of considering these higher-level factors in the analysis to capture the full dynamics affecting electoral success.

The coefficient of quota compliance becomes statistically significant at the 10 percent level when we account for gender in the entire sample. However, this result should be interpreted with caution. It suggests that the treatment effect might be predominantly driven by the stronger influence of men, rather than accurately reflecting the true impact of the treatment. This interpretation is further supported by the negative coefficient associated with gender (0 for men and 1 for women) and the relatively unchanged coefficients for other variables. Additionally, the random effects of party and province remain largely consistent in terms of standard deviations and R^2 when compared to model 1. These observations reinforce our argument regarding the potential variability of the predictors’ effects in explaining the outcome, emphasizing the need for caution when analyzing treatment effects in this context.

Lastly, one might wonder to what extent adherence to quota rules influences election outcomes when a standard logistic regression is employed. We fit this model in Table 3 using the initial, pre-matching data. We run the same model as before, dividing the analysis into combined and female-only samples. The results overall remain consistent, showing null effects of the treatment, except for logit 4 when we account for the gender factor. Additionally, compared to the propensity score analyses, the standard model tends to slightly overestimate the influence of control variables, possibly due to the potential imbalance of covariates within each of the predictors. This highlights the limitations of the standard logistic regression in capturing the ‘true’ effects, emphasizing the need for more refined techniques such as PSM.

All in all, the null results of the treatment remain consistent across different model specifications and sample sizes. This consistency allows us to claim that the level of women’s representation across provinces in Indonesia is not driven by high compliance with the 30 percent gender quota policy.

Table 3: Standard logistic regression (naive results)

	logit 1	logit 2	logit 3	logit 4
(Intercept)	−3.298***	−1.257***	−2.561***	−1.618***
	(0.139)	(0.252)	(0.051)	(0.101)
Treatment	0.067	0.028	0.004	0.102+
	(0.148)	(0.155)	(0.056)	(0.059)
List position		−1.008***		−0.603***
		(0.047)		(0.016)
District magnitude		0.127***		0.119***
		(0.022)		(0.010)
Gender				−0.593***
				(0.057)
Num.Obs.	11 721	11 721	32 185	32 185
AIC	3768.5	2892.4	16 644.8	13 747.3
BIC	3783.2	2921.9	16 661.6	13 789.2
Log.Lik.	−1882.255	−1442.188	−8320.402	−6868.660

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Discussion