

# PS531 Pre-Analysis Plan

Zach Salazar

7/17/23

## **1 Describe a substantive question in social science. What theory are you assessing? Why should anyone care? (2 paragraphs)**

Does past military experience, specifically military rank, affect foreign policy decision-making? Current research on the impact of past military experience in leaders has only examined experience with and without combat (Horowitz & Stam, 2014). While the results were significant, I believe we can go deeper to truly understand the impact of military experience by looking at the command structure of the military, specifically focusing on the differences between officers and enlisted personnel.

Leader studies in Political Science have gained traction in recent years, with scholars examining aspects such as achievement motivation (Winter, 2010), rebel experience (Fuhrmann & Horowitz, 2015), and resolve (Kertzer, 2016), among others. This paper aims to expand that area, specifically building upon work from Horowitz and Stam (2014) about prior military experience. While this paper is primarily International Relations (IR) focused, the theory being assessed is also useful for American politics, as the decision-making process and background of a leader is important during presidential elections.

## **2 The study you propose involves learning about a theory by observing certain of its implications. What one or two hypotheses that arise from the theory are you planning to assess? Why or how does the theory justify your expectations about these hypotheses? (1 or 2 paragraphs)**

In order to answer the question of how military rank influences foreign policy, we must first examine the aspects of foreign policy that are most impacted by a leader's decision-making.

**Hypothesis 1:** *Leaders with higher military ranks will initiate less disputes.*

I believe leaders who attained higher ranks will be more hesitant to start disputes due to having a better understanding of war. Similar to how combat experience makes leaders more hesitant to use force (Horowitz & Stam, 2014), I believe higher ranking officers will exhibit similar behavior. However, I will need to be mindful of combat experience becoming a confounding variable, as most leaders with higher ranks will also have combat experience.

**Hypothesis 2:** *Leaders with higher military ranks will win more disputes.*

While I believe the higher ranked leaders will be less likely to start disputes, I also believe they are more likely to win due to having more military experience. The higher rank will also manifest itself in better military strategy, with another hypothesis possibly being less fatalities incurred by the leader's side. However, there is little research concerning fatalities and foreign policy as it relates to political psychology.

### **3 What data and research design will help you answer this question? Why are you making these choices? (Remember that a statistical model is not a research design.) (2 paragraphs)**

To answer my research question, I plan to use an observational study, primarily utilizing the Militarized Interstate Dispute (MID) and the Leader Experience, Attribute, and Decision (LEAD) datasets. The MID dataset contains data for every conflict from fishing disputes to total wars from 1816-2014 (Maoz et al., 2019). The LEAD dataset contains characteristics about every world leader from 1875-2004 (Ellis et al., 2015). For military rank, the Veteran's Association (VA) has the military ranks for every president available on their website. To organize this for analysis, I assigned a number for military rank, or lack thereof (see Table 1). This allows me to expand beyond the limitations of the LEAD dataset and include presidents going back to James Madison, as this is the earliest president to have MID data available.

To measure my control variable of relative power, I will also use the National Material Capabilities (NMC) dataset. This dataset is organized by year and gives a decimal value, representing the percentage of military power a state has relative to the world out of 100%. For example, in 1950, the United States has a CINC score of 0.28444302, meaning they had 28.4% of the world's military power during that year. For each president, they are given a CINC score that is the average of their years in office. This is a useful control variable since military power could influence foreign policy significantly.

#### **4 What are the advantages and disadvantages of this research design to addressing your question? (2 paragraphs discussing both advantages and disadvantages of the research design; could be 1 paragraph for advantages and 1 for disadvantages or combined discussion across 2 paragraphs.)**

The main advantage of this research design is that it is a longitudinal study. This allows me to observe changes at both the individual and group levels. For example, there could be trends specific to leaders of autocratic countries that are not present in democracies. This would open up new research opportunities to expand this research beyond great powers. It does all of this over a period of nearly 200 years, which is a significantly longer time frame than an experiment would allow. Finally, since this design has the potential for greater external validity due to the real-world implications of this research.

The primary disadvantage of this research design is the difficulty in proving causation. As mentioned previously, confounding variables like combat experience and relative power could be driving any relationship I find. In the case of combat experience, it has already been found to be significant in predicting aggression in foreign policy (Horowitz & Stam, 2014). On top of this, there is the overarching issue of the difficulty in proving leaders matter (Jervis, 2013).

#### **5 Describe your measures and any indices you construct. (1 paragraph)**

For military rank, I have the ranks ordered starting with no experience, followed by enlisted<sup>1</sup>, then all ten officer ranks. As this project expands to include great powers, I intend to normalize all ranks to NATO ranks for easier comparison between countries. For countries where there is less structure in the military, this organization will be difficult to implement. While enlisted ranks will remain relatively static throughout the world, officer ranks and their responsibilities might differ greatly. A possible solution is grouping officer ranks together if they are similar, which would make comparison to NATO ranks easier.

---

<sup>1</sup>Enlisted ranks refer to non-officer personnel.

Table 1

Rank	Number
None	0
Enlisted	1
OF-1	2
OF-2	3
OF-3	4
OF-4	5
OF-5	6
OF-6	7
OF-7	8
OF-8	9
OF-9	10
OF-10	11

This ordering is subject to change as I learn more about the intricacies of officer ranks, as the experiences of an OF-1 and OF-2 might not differ enough to require separate classifications.

## **6 Use data to make the case that your research design allows you to interpret observed quantities (like observed data comparisons or parameters of models fit to data) as theoretically relevant and clear.**

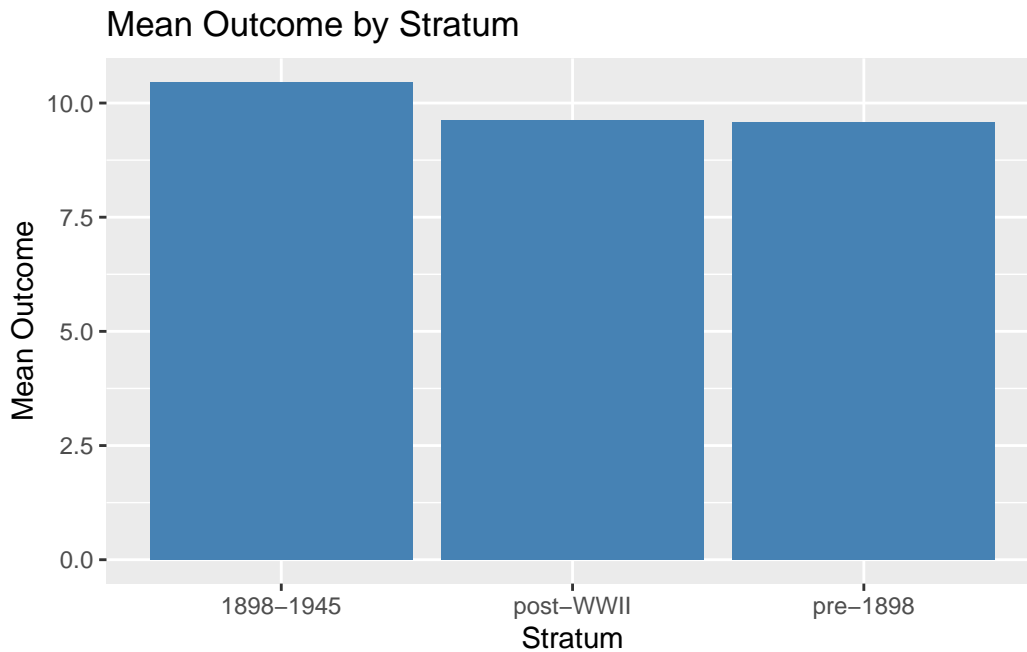
**6.1 If you are using an observational study design then explain how you will make the case for interpretable comparisons (this is the same as question as ‘What is your identification strategy?’). That is, explain how you will use statistical adjustment (like matching or covariance adjustment aka “controlling for”) to persuade yourself and others that the comparison that you are showing reflects what you say it does. If you are making comparative or causal inference, I assume you will explain the natural or quasi-experimental design and approach you will be using here. “I controlled for a lot of background variables in a linear model.” will not be acceptable here. If you are making population inferences, you should explain your approach as well. (2 paragraphs plus some tables or figures)**

In some leader studies in IR that focus on US presidents, a common cutoff point is 1898 due to the US becoming a world power that year(Harden, 2021). Others will only examine post-World War II leaders due to the radical change in the international system following the war

(Yarhi-Milo, 2018). Many different groups can be created for the leaders based on the time they were in office. This would allow for more consistent measurement as the leaders would all be in similar international environments.

For these reasons, stratification will be the best method of adjustment for my project. The example below shows a stratification table where three groups are created for the time periods most commonly used in leader studies. While this will not account for other potential confounding variables like combat experience or relative power, it will at least help control for most systemic variables.

```
# A tibble: 3 x 2
  stratum mean_outcome
  <chr>      <dbl>
1 1898-1945    10.5
2 post-WWII     9.62
3 pre-1898     9.59
```



**6.2 If you are using an observational study design, explain how you will judge the success of your adjustment strategy. For example, you may explain here about balance tests and other diagnostics that refer to the problem of adjustment for confounding or making the case for an as-if-randomized comparison, or an as-if-randomly sampled set of observations, etc.. (1 paragraph)**

To test the performance of my stratification, I will perform balance tests. Specifically, I will calculate summary statistics between each strata. If the groups have significantly different means for a particular covariate, it suggests that the stratification adjustment did not effectively balance that covariate. Randomization would not be ideal due to the small sample size. Even if the project expands to include great powers or all countries, I still would not use randomization because all of the leaders were specifically chosen and classified by my own set of criteria.

**7 Explain your plans for any missing data or extreme outcome or covariate values you may encounter when you get the real data (or perhaps you have the background data but not the real outcomes, so you can explain your plans for such data issues in that case here too). (1 or 2 paragraphs)**

This project will eventually expand to include at least all other great powers leaders<sup>2</sup>, possibly expanding to the leaders of every country if the data is easily accessible. When I began working on this project, finding data for leaders of other countries was difficult. For example, France has at least five leaders where I can not find any information about their military rank, despite knowing they served. For these cases, I intend to perform the analysis without these leaders included. For most of the leaders where little information is available, they are usually inconsequential, having only been in office for a few months or less. Most of these leaders have no MIDs attributed to them, meaning there is little harm in leaving them out.

For extreme values, this is something that I have already begun dealing with in my preliminary analysis of the great powers. Leaders like Hitler and Stalin have significantly more MIDs due to being heavily aggressive and being in power for decades, respectively. To deal with this problem, changing from raw counts to averages could lessen the impact of the extreme outliers. This is already a small issue among US Presidents, as some were in office for 8 years or more<sup>3</sup> while others were in office for 4 years or less, giving them substantially lower total numbers for variables like MID initiations.

---

<sup>2</sup>Great powers include Russia, China, Great Britain, Germany, France, and Italy.

<sup>3</sup>Franklin Roosevelt was president for 12 years.

**8 What statistical tests do you plan to use? Explain why you chose these tests and any decision making criteria you will use upon seeing the results of the tests. You should also engage with the problem of multiple testing here if you are going to show the results of more than one test. (Recall that confidence intervals and hypothesis tests convey more or less the same information. So a confidence interval is a form of testing.) (1 paragraph)**

I intend to use multivariate regression analysis to test my hypotheses because my research design has multiple dependent variables. While other methods can be used, this will test my hypotheses in the most simple way, while allowing me to expand the project to include new methods if they are necessary. When I started this project a year ago at Ohio State, a professor recommended using a negative binomial due to my data being over dispersed. However, during my time in Jake's class, I began to see the issues that can arise when using a negative binomial, as evidenced by the problems I encountered during the explorations.

**9 Explain how you will judge the performance of those tests. Will you only use simple false positive rate and power? Or do you need to add family-wise error rate? false discovery rate? Or something else? Explain why you made this choice. (1 paragraph)**

To judge the performance of my tests, I will use FWER. Since I am looking at more than one dependent variable, there is the possibility of a Type I error occurring across the multiple tests I will conduct. Currently, my dataset contains information about initiations, conflict duration, conflict outcomes, and fatalities incurred. These all cover different areas of foreign policy which makes Type I errors more likely.

**10 Show and explain how your test performs in regards those properties (at least you will show false positive rate and power). (2–4 paragraphs)**

Using the Bonferroni correction to control for FWER is ideal since I plan to examine multiple aspects of foreign policy, which is calculated as follows:

$$\alpha_{\text{adj}} = \frac{\alpha}{k}$$

This method is ideal because it allows flexibility in adjusting significance levels. I will also be able to customize the stringency of the correction to strike a balance between controlling errors and maintaining statistical power.

However, the Bonferroni correction can be too conservative, which means it can be overly stringent. This increases the likelihood of Type II errors, which is caused by the assumption that the tests are independent. A possible solution is using a more powerful method like the Šidák correction. However, this may become less of an issue if the sample size increases after introducing more countries to the study. Further testing will be needed to determine the best method for my data.

## **11 What statistical estimators do you plan to use? Explain why you chose these estimators. Especially explain what is your target of estimation — what is the estimand? (1 paragraph)**

I intend to use Difference-in-Differences (DiD). This method helps to control for time-invariant unobservable factors and provides a more robust estimation of the causal effect. The target of estimation is the causal effect of past military ranks of presidents on their foreign policy decision-making. By estimating this effect, I aim to quantify the relationship between past military ranks and foreign policy decision-making while accounting for other relevant factors.

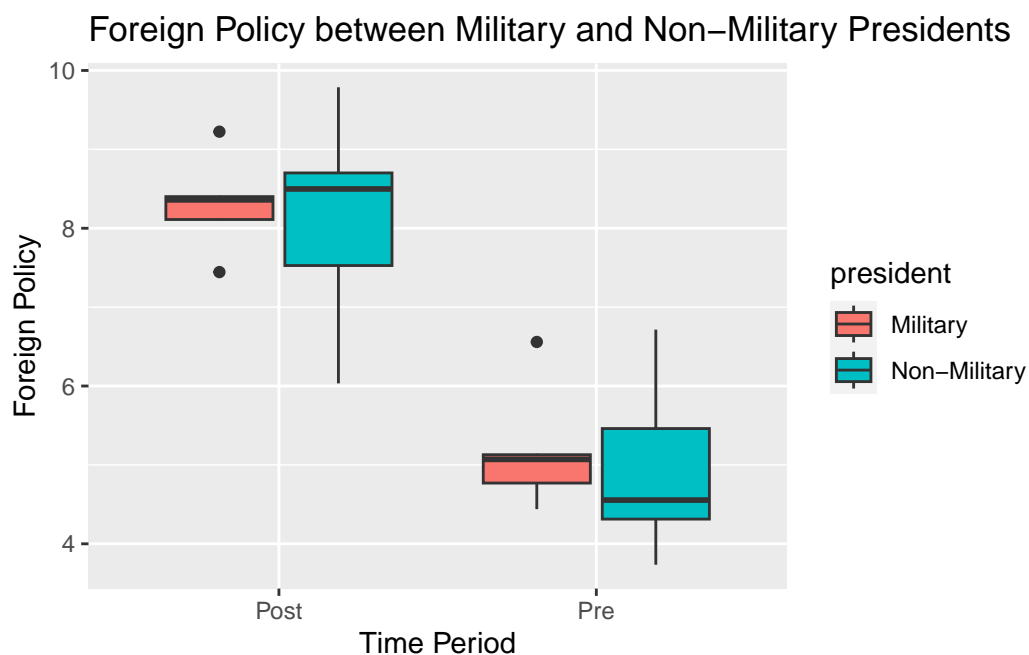
## **12 Explain how you will judge the performance of those estimators (especially bias and MSE)? (1 paragraph)**

I can calculate the MSE by comparing the predicted values from my DiD model to the actual values of the outcome variable. To test bias, I will need to consider factors such as unobserved confounders and time-varying factors that affect both the treatment and outcome. For MSE, I can compare the predicted values from my DiD model to the actual values of the outcome variable. Additionally, I can consider conducting cross-validation or bootstrapping techniques to obtain more robust estimates of the MSE.

## **13 Show and explain how your estimator performs in regards those properties (at least bias and MSE). (2–4 paragraphs)**

Below is an example of DiD using fake data. This is slightly different than my own data as this only looks at military and non-military experience. However, the process will be the same for the real data.





Call:

```
lm(formula = foreign_policy ~ time * president, data = data)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.07596	-0.59723	-0.09367	0.52693	1.75938

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	8.30790	0.46965	17.690	6.29e-12 ***
timePre	-3.11433	0.66418	-4.689	0.000246 ***
presidentNon-Military	-0.19856	0.66418	-0.299	0.768825
timePre:presidentNon-Military	-0.03933	0.93930	-0.042	0.967119

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.05 on 16 degrees of freedom

Multiple R-squared: 0.7366, Adjusted R-squared: 0.6872

F-statistic: 14.92 on 3 and 16 DF, p-value: 6.766e-05

```
[1] "Bias: NA"
```

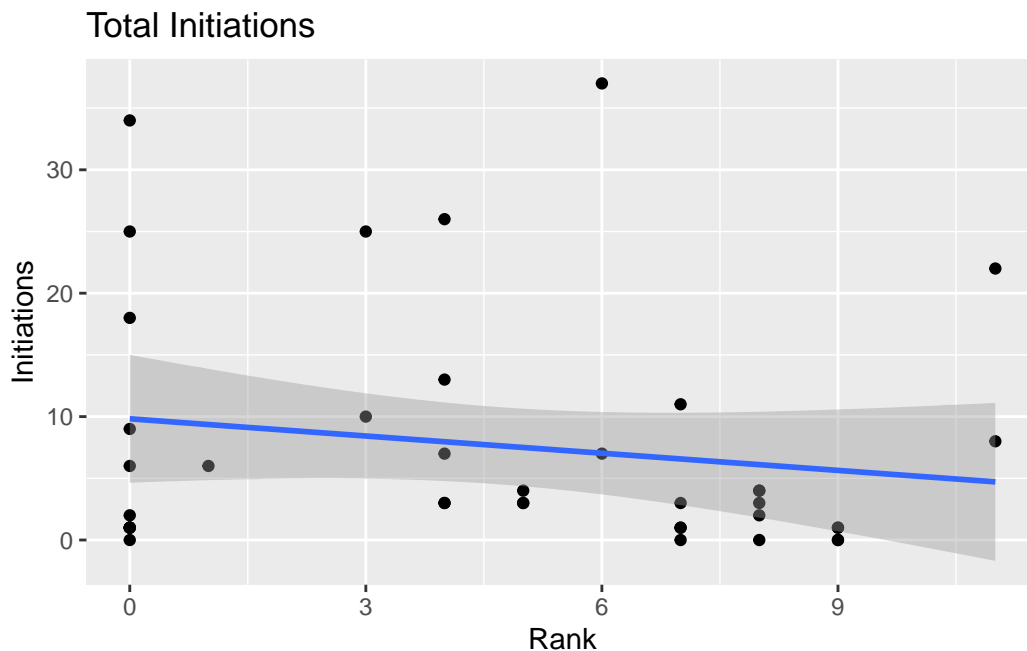
```
[1] "MSE: 0.882283025085199"
```

As shown by the results, the bias is given a value of NA. From what I could find, this is either an error, or there is not enough variance in the fake data. For MSE, we are given a value of 0.88. In this case, an MSE of 0.88 suggests that, on average, the predicted values from the DiD model differ from the actual values by approximately 0.88 units squared.

**14 Make one mock figure or table of the kind you plan to make when you use the actual outcome. Interpret the results of the mock analysis as if it were the real analysis. Saying something like, “If the real outcome were as I have simulated it, then the following table/figure would mean such and so about the theory.” (1 paragraph)**

The plot presented below (Figure 1), which is a linear regression plot, shows that as military rank increases, the amount of MID initiations decreases slightly. This is consistent with my hypothesis that presidents who attained higher ranks in the military are more selective with starting conflicts. While the outcome is not as strong as I anticipated, this gives me hope the results will continue in this direction when I expand the data to great powers and possibly every country. However, this is only one aspect of foreign policy, and further analysis will be needed on aspects such as conflict outcomes. Further plots and tables can be found in the appendix.

**Figure 1**



**15 Include a code appendix and a link to the github repository for this paper.**

<https://github.com/zachsalazar/531-pre-analysis-plan>

## References

- Ellis, C. M., Horowitz, M. C., & Stam, A. C. (2015). Introducing the LEAD data set. *International Interactions*, 41(4), 718–741.
- Fuhrmann, M., & Horowitz, M. C. (2015). When leaders matter: Rebel experience and nuclear proliferation. *The Journal of Politics*, 77(1), 72–87.
- Harden, J. P. (2021). All the world’s a stage: US presidential narcissism and international conflict. *International Studies Quarterly*, 65(3), 825–837.
- Horowitz, M. C., & Stam, A. C. (2014). How prior military experience influences the future militarized behavior of leaders. *International Organization*, 68(3), 527–559.
- Jervis, R. (2013). Do leaders matter and how would we know? *Security Studies*, 22(2), 153–179.
- Kertzer, J. D. (2016). *Resolve in international politics*. Princeton University Press.
- Maoz, Z., Johnson, P. L., Kaplan, J., Ogunkoya, F., & Shreve, A. P. (2019). The dyadic militarized interstate disputes (MIDs) dataset version 3.0: Logic, characteristics, and comparisons to alternative datasets. *Journal of Conflict Resolution*, 63(3), 811–835.
- Winter, D. G. (2010). Why achievement motivation predicts success in business but failure in politics: The importance of personal control. *Journal of Personality*, 78(6), 1637–1668.
- Yarhi-Milo, K. (2018). *Who fights for reputation: The psychology of leaders in international conflict*. Princeton University Press.