

Accepted Manuscript

DOI: <https://doi.org/10.1177/00223433231201447>

Citation: Bertoli, A., Dafoe, A., & Trager, R. (2024). Leader age and international conflict: A regression discontinuity analysis. *Journal of Peace Research*, 61(4), 643-658.

This article has been accepted for publication and has undergone full peer review. However, this version does not have the copyediting, typesetting, pagination, and proofreading processes, which may result in differences between this version and the final Version of Record.

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information, please visit the publisher's website.

Leader age and international conflict: A regression discontinuity analysis

Andrew Bertoli, School of Global and Public Affairs, IE University

Allan Dafoe, Centre for the Governance of AI

Robert Trager, Department of Political Science, UCLA

Abstract: Does leader age matter for the likelihood of interstate conflict? Many studies in biology, psychology, and physiology have found that aggression tends to decline with age throughout the adult lifespan, particularly in males. Moreover, a number of major international conflicts have been attributed to young leaders, including the conquests of Alexander the Great and the ambitious military campaigns of Napoleon. However, the exact nature of the relationship between leader age and international conflict has been difficult to study because of the endogeneity problem. Leaders do not come to power randomly. Rather, many domestic and international factors influence who becomes the leader of a country, and some of these factors could correlate with the chances of interstate conflict. For instance, wary democratic publics might favor older leaders when future international conflict seems likely, inducing a relationship between older leaders and interstate conflict. This article overcomes such confounding by using a regression discontinuity design. Specifically, it looks at close elections of national leaders who had large differences in age. It finds that when older candidates barely defeated younger ones, countries were much less likely to engage in military conflict. Its sample is also fairly representative of democracies more broadly, meaning that the findings likely hold true for cases outside the sample. The results demonstrate the important role that individuals play in shaping world politics. They also illustrate the value of design-based inference for learning about important questions in the study of international relations and peace science.

Keywords: leaders, international security, regression discontinuity, natural experiment, political psychology, elections

Corresponding Author: abertoli@faculty.ie.edu

In what ways are the foreign policies of states shaped by the individuals who lead them? This question has played a central role in the study of interstate politics going back to Thucydides, who devoted great attention to leader qualities and their apparent impact on state behavior (Connor, 1984). This focus is perhaps most evident in the contrast between the young, bold Alcibiades, who favored war with Syracuse, and the older, cautious Nicias, who opposed it.

In this article, we take up the question of whether leader age matters for the likelihood of international conflict. Focusing on democracies, we employ a regression discontinuity design that takes advantage of the as-if random outcomes of very close elections between potential leaders with large differences in age. Our results indicate that electing older leaders does make countries less likely to act aggressively towards other states. We also find that the estimated effect is much larger when presidential democracies have united governments, possibly because leaders exercise greater influence over foreign policy when their co-partisans control the legislative branch (Howell & Pevehouse, 2005, 2007). In addition, our estimates are much larger for more consolidated democracies, suggesting that leader traits matter more when democratic leaders have greater political security (Buono de Mesquita et al., 1999; Chiozza & Goemans, 2003; Palmer, London & Regan, 2004).

These findings accord with many recent studies that contain a wealth of evidence about the effects of leaders on international conflict (e.g. Saunders, 2011; Weeks, 2014; Barceló, 2020; Dube & Harish, 2020). Our regression discontinuity design overcomes a key challenge that makes research on this topic difficult—the endogeneity problem. Put simply, who comes to power in a country is not random. It is influenced by many factors at the domestic and international levels, and some of these factors could correlate with the chances of interstate conflict. This can make it hard to know whether differences in state behavior

result from leaders themselves or from the broader circumstances that facilitated their rise to power. Because we focus on close elections where who became leader was essentially as-if random, our estimates should be largely free of such selection effects.

Our results also fit with the well-established finding in other fields that aggression tends to decline over the adult lifespan, particularly in males (Birditt & Fingerman, 2005; Williams et al., 2006; Neupert, Almeida & Charles, 2007). Moreover, this conclusion receives further support from the many international confrontations that have been attributed to the desire of young leaders to overturn the status quo. These cases range from the wars of Alcibiades and Alexander the Great to the 23-year-old Austro-Hungarian Emperor Franz Josef's break with Russia during the Crimean War, a decision he made against the strong advice of his 80-year-old elder statesman, Prince Metternich (Rich, 1985; Trager, 2012). This disastrous decision that the Emperor made as a young man put him on the path to fight the First World War at the age of 83 because that early blunder precipitated the security competition with Russia in the Balkans.

This article contributes to political science in several important ways. First, the results provide strong evidence from a natural experiment that leaders do have a significant independent impact on foreign policy. Recent debates about medical research have made clear the great value of research designs that have random or as-if random treatment assignments for drawing causal inferences. Such opportunities are rare in the study of international conflict, but when they occur they can provide very valuable insights into the factors that drive state behavior (Jones & Olken, 2009; Dube & Harish, 2020). We thus believe that our article makes a very important contribution in this regard. Our results emphasize the importance of individual leader traits as drivers of international outcomes and should encourage more research in this area.

Second, our findings build a bridge between international relations scholarship and a much broader body of research about age and aggression that spans the fields of biology, psychology, physiology, and criminology. This is important because one of the primary goals of science is to create unified bodies of knowledge, not just within research programs but also between them. Our results thus help create connections between scholarly fields that can improve our understanding of human behavior more broadly.

Third, our findings affirm the theory that youth may partly explain why some notable leaders of the past have launched aggressive military campaigns. For example, the wars of Alexander the Great, Hannibal, Charlemagne, Suleiman I, Gustavus Adolphus, Frederick the Great, Napoleon, and Shaka may be partly explained by how young these men were when they came to power. Our findings should encourage historians and social scientists to take leader age very seriously when investigating the causes of international conflicts.

Age and aggression

The link between age and aggression has been well-documented across many different cultural contexts (Daly & Wilson, 1990; Steffensmeier & Streifel, 1991; Sweeten, Piquero & Steinberg, 2013). Psychologists have demonstrated that older adults adopt less aggressive strategies in interpersonal relationships (Neupert, Almeida & Charles, 2007; Hay & Diehl, 2011). All else equal, they are less likely to yell or name call and more likely to endorse withdrawing from conflictual situations.

Similarly, criminologists have described an ‘age-crime curve’, according to which criminal behaviors rapidly decline from late adolescence through young adulthood and continue decreasing through old age (Farrington, 1986; Steffensmeier et al., 1989; MacLeod, Grove & Farrington, 2012; Nyseth Brehm, Uggen & Jean-Damascène, 2016). The decline

in crime with age seems to apply to both violent and non-violent types of crime (Steffensmeier et al., 1989), and it continues as people in their 40s and 50s progress to their 60s and 70s (Steffensmeier et al., 1989; MacLeod, Grove & Farrington, 2012).

Some explanations of these changes focus on life-stage or role changes associated with aging, such as entry into marriage and raising children, which are associated with declines in crime rates. It is likely that such role changes are only part of the explanation, however. A principle reason for this is that even within sub-populations that experience role changes, the incidence of violent behaviors tends to decline with age (Siennick & Osgood, 2008). Existing studies of role changes also have not employed causal identification strategies, and thus it is possible that other age-related changes drive these associations.

Psychologists have identified several psychological and physiological processes associated with aging that may relate to declines in aggressive behaviors. We group the most significant findings into four areas: (1) changes in emotion regulation, (2) testosterone decline, (3) changes in attention and focus, and (4) changes in amygdala activity. First, older adults tend to be better at regulating their emotional states after negative or adverse experiences. They move out of negative emotional states more quickly and are less reactive to interpersonal stressors. In part, but only in part, this is because they employ less inflammatory interpersonal conflict strategies (Mather, 2012). The shorter time horizons of adults in the later stages of their lives may also lead them to prioritize well-being in the moment and therefore allocate more cognitive and social resources to the regulation of emotion (Carstensen, Fung & Charles, 2003; Carstensen, 2006).

Second, because testosterone levels decline predictably and continuously in men after the mid 20s (Fabbri et al., 2016), they offer another plausible explanation for the age-related changes in aggressive behaviors. However, the exact nature of the link between

testosterone and aggression in men is highly contested (Eisenegger et al., 2010; van Honk et al., 2012). The most convincing evidence of a causal relationship between testosterone and aggression comes from experimental animal studies. These show that aggression can be induced by manipulating testosterone levels (Edwards, 1969; Albert et al., 1990). Studies of humans show correlations between testosterone and aggression, but scholars disagree about whether the relationship is causal (Berman, Gladue & Taylor, 1993; Book, Starzyk & Quinsey, 2001). Nevertheless, the extensive research that has been conducted on this subject does provide some support for the hypothesis that testosterone increases aggression and that declines in testosterone levels with age may be a factor leading to the decline in aggressive behaviors.

Third, older adults develop the ability to focus on positive over negative information and be less easily influenced by negative stimuli (Williams et al., 2006; Kisley, Wood & Burrows, 2007; Marquez-González et al., 2008; Mather, 2012). Relative to younger adults, they tend to look away from negative images (Knight et al., 2007). This tendency to avoid the negative is *reinforced* when they are induced into a negative mood, whereas younger adults focus on the negative when in a negative frame (Isaacowitz et al., 2008). Moreover, older adults are less likely to be distracted by negative stimuli (in contrast to positive stimuli), and stressful recent events are less predictive of cognitive interference (Stawski, Mogle & Sliwinski, 2011). Older adults also show higher relative activity in the region of the brain associated with emotion regulation, the prefrontal cortex, in response to negative stimuli, an indication that the differences in response to negative stimuli are driven by differential recruitment of brain resources (Fischer et al., 2005).

Fourth, younger and older adults show different amygdala responses to positive and negative stimuli. The amygdala has been shown to play a fundamental role in emotional decision-making, including in aggressive responses to stimuli. Heightened amygdala activity is associated with aggressive behaviors (Siever, 2008). While younger adults show the highest amygdala responses to negative stimuli, older adults show the highest responses to positive stimuli and much lower amygdala responses to negative stimuli (Mather et al., 2004; Leclerc & Kensinger, 2011). These differences do not appear to be driven by declines in amygdala function with age (Mather, 2012). Thus, in response to negative stimuli, older adults show both a relative increase in the activity of the prefrontal cortex, which regulates and moderates emotional responses, and a decrease in the activity of the amygdala, which drives aggressive responses.

In sum, the reason why older people tend to behave less aggressively than younger ones still remains uncertain. Several plausible mechanisms have been proposed, and it is likely that some of them reinforce each other. Nevertheless, the research discussed in this section has established a clear link between age and aggression that has appeared across a wide range of contexts.

Given this link, we might expect countries to behave more peacefully when they are ruled by older leaders compared to younger ones. However, there are reasons to question whether the relationship between age and aggression will appear at the international level as well. We discuss these possibilities in the next section.

Are leaders different?

Do world politics operate in a fundamentally different way than the other domains in which age and aggression have been studied? There are several reasons to suspect that the answer may be yes. These are:

- (1) How people tend to behave in interpersonal situations might have little bearing on their foreign policy stances.
- (2) People who rise to the level of potentially becoming national leaders go through a selection process that might make them similar to each other.
- (3) Domestic and international factors may largely constrain leader behavior, muting the effects of leaders' personalities.
- (4) Older leaders may have more consolidated political power, giving them more freedom to take military action abroad.
- (5) Older leaders may have shorter time horizons and therefore be in a greater rush to establish their legacies.
- (6) Older leaders may have more confidence in their abilities to manage international crises and conflict.

Regarding the idea that (1) how people tend to behave in interpersonal situations might not carry over into their foreign policy stances, recent research indicates that such spillover does happen. For example, in a study on 412 individuals in the United States, Kalmoe (2013) found a strong association between how people scored on a psychological aggression test and support for various forms of state violence, including using military force against another country. McDermott & Hatemi (2017) found similar results for a sample

of 586 people in Australia. These findings support the conclusions of prior studies on college students (Anderson et al., 2006; McIntyre et al., 2007). In short, how people tend to behave in interpersonal situations does seem to predict their foreign policy stances.

What about the possibility that (2) selection processes may make politicians who rise to the level of potentially becoming national leaders very similar to each other? If this were true, then younger and older leaders might have very similar personalities. No doubt, selection processes do influence who comes to power in a country. For example, more intelligent and more ambitious people are more likely to rise to the level of national leader. However, if these selection processes were such a powerful source of conformity, then the leaders of a given country should tend to look very similar to each other, especially after leadership transitions where little about the country changed besides the leader. Yet many examples show that the personality traits and foreign policy approaches of outgoing leaders often differ greatly from new leaders. Such cases include Ancient Rome's transition from the stoic philosopher Marcus Aurelius to the violent sociopath Commodus, the Soviet Union's transition from the hawkish Konstantin Chernenko to the much more dovish Mikhail Gorbachev, and China's transition from the domestically-focused Hua Guofeng to the much more internationally assertive Deng Xiaoping. In democracies, recent elections similarly demonstrate that candidates for national office can have very large personality differences and foreign policy approaches, such as Donald Trump and Joe Biden. Thus, selection processes do not seem to homogenize leaders to a very large extent.

The question of whether (3) international and domestic factors greatly constrain leader behavior has been one of the most important debates in the study of international relations. Some prominent scholars have argued that strategic realities greatly limit what leaders can do in world politics (e.g. Waltz, 1979; Mearsheimer & Walt, 2003). For example,

Mearsheimer (2014: 10) recently asserted, ‘it does not matter much who is in charge in Cairo or Damascus. The United States has a rich history of working with leaders of all types’. Nonetheless, a wave of recent studies find that certain types of leaders do seem to increase the likelihood of international conflict (e.g. Byman & Pollack, 2001; Jones & Olken, 2009; Weeks, 2014; Horowitz et al., 2018; Barceló, 2020; Dube & Harish, 2020). Most of these studies have either been qualitative or large-*N* cross-national regression analyses. This article analyzes a natural experiment that finds further evidence that leaders do have an important independent impact on foreign policy.

Possibilities (4)-(6) were proposed in two previous studies about leader age and international conflict (Horowitz, McDermott & Stam, 2005; Horowitz, Stam & Ellis, 2015). These studies made very important contributions to the field by drawing attention to the importance of leaders and particularly the potential relationship between leader age and interstate violence. Given the well-established finding from other fields that aggression tends to decline with age, as well as the many famous historical examples of aggressive young leaders, the possible link between leader age and interstate conflict was definitely worth exploring. These two studies were also noteworthy because they discovered a very surprising correlation. They found that older leaders tend to engage in military conflict at much higher rates than younger leaders. To our knowledge, no study in the broad literature on age and aggression in other fields has found that aggressive tendencies tend to increase over the adult lifespan. However, these two studies came to exactly that conclusion for leaders. Also surprisingly, their results were driven by democracies rather than non-democracies.

We will first discuss the theoretical reasons the authors used to explain these surprising findings, and then we will offer an alternative interpretation of their results. Regarding the

idea that (4) older leaders may have more consolidated political power, it is not clear why older leaders would use their political capital to engage in conflict as opposed to pursue more cooperative policies that encourage international peace. In other words, why would increased freedom in foreign policy make a leader more likely to start conflicts rather than to sign treaties and negotiate with other countries to resolve disputes peacefully? There may be some causal logic for why greater autonomy would make it easier for leaders to take aggressive foreign policy moves as opposed to more conciliatory ones. However, it is not clear why this would necessarily be true. It seems that increased leader autonomy could encourage either international peace or conflict, depending on what the leader's goals were.

Regarding the idea that (5) older leaders may have shorter time horizons, we do not think that time horizons can easily explain why older leaders would behave more aggressively. This is because there is a difference between an individual's biological time horizons and a national leader's political time horizons. While older individuals may have shorter biological time horizons, older leaders do not necessarily have shorter political time horizons. In fact, in the data from the 2015 study, which includes more than 2,000 leaders from 1875-2001, older leaders tended to stay in power longer than younger leaders. In democracies, leaders who took office at about age 30 had tenures that lasted around 1.6 years on average, whereas leaders who took office around age 70 had tenures that lasted for about 4.0 years on average. Likewise, in non-democracies, leaders who took office at about age 30 had tenures that lasted around 5.4 years on average, whereas leaders who took office around age 70 had tenures that lasted for about 6.6 years on average. Therefore, depending on whether one is referring to biological or political time horizons, the time-horizons mechanism seems to lead to conflicting predictions.

Finally, the idea that (6) older leaders may have more confidence in their abilities to manage international conflict is also questionable. They may have more experience, but that experience could teach them humility rather than hubris. Past research on how confidence in performing cognitive tasks changes with age comes to mixed results (Touron & Hertzog, 2004; Bruine de Bruin, Parker & Fischhoff, 2012). Moreover, even if older leaders do tend to have more confidence than younger ones, it would not necessarily make them more likely to engage in conflict. Rather, it might make them more confident in their ability to reach favorable settlements without fighting. Their additional experience over younger leaders could help them in this regard.

Instead of older leaders' political capital, time horizons, or confidence, there is an alternative reason why they may engage in interstate conflict more often than younger ones. When times are turbulent, democratic publics would likely prefer older candidates, as we saw in the 2020 Democratic primary. Biden and Sanders flourished while younger, less experienced candidates struggled to gain support. Another example is Winston Churchill, who was appointed Prime Minister of Great Britain shortly after the start of the Second World War. Then, despite overwhelming popularity as a result of his successful prosecution of the conflict, he was voted out of office shortly after the Allied victory in Europe. The British people judged him the best leader to fight the war, but not the best candidate to lead Britain in peacetime (Jenkins, 2001). He was succeeded by the younger politician Clement Attlee. Thus, older democratic leaders may be much more likely to come to power in moments of crisis and instability. This political selection effect could explain the strong empirical link between older leaders and interstate conflict.

In contrast to the two prior studies, our article tests whether *having* older leaders makes countries more or less likely to engage in interstate conflict. Put another way, we examine

whether older or younger leaders are more likely to engage in interstate conflict, controlling for environmental factors at the domestic and international levels through regression discontinuity. Importantly, we do not dispute the strong statistical association between older leaders and interstate conflict found in past research. In fact, this surprising discovery served as an inspiration for our research. It is important to note, however, that neither of the two past studies attempted to control for environmental factors at both the domestic and international levels. The 2005 study only controlled for international factors and regime type, while the 2015 study controlled mainly for leader traits. The quantitative models in these studies therefore uncovered important statistical associations, but they did not directly address the main question that we ask in this article. They thus might be thought of as adhering to the regularity or mechanistic versions of causality rather than the counterfactual framework (Box-Steffensmeier, Brady & Collier, 2008).

Our approach is rooted in the counterfactual version of causality. We are interested in how countries' foreign policies might differ if they had older vs. younger leaders. This type of question can best be answered with a natural experiment, which controls for observable and unobservable factors through the as-if random assignment of treatment (Box-Steffensmeier, Brady & Collier, 2008; Robinson, McNulty & Krasno, 2009; Imai, 2017). When comparing older and younger leaders, the natural experiment approach is particularly essential because we lack an accurate measure of the key potential confounder—general levels of instability. It would be difficult to operationalize this variable, and any measurement error in coding it could bias the results. The best way to control for this variable is through the as-if randomness of treatment assignment. We outline our research approach in the next section.

Research design

To test whether younger or older leaders are more prone to engage in military conflict (controlling for environment), we employ a regression discontinuity design. Specifically, we look at close elections of national leaders from 1815-2014. Following our pre-analysis plan, we focus on elections where the top two candidates were at least ten years apart in age. This approach gives us a dataset where the average difference in age between the younger and older leaders was around 18 years. Thus, our cases provide us with fairly substantial variation in our independent variable.

At the same time, there were only three female candidates in our close elections, which means that our analysis is largely limited to the male gender. This is an important caveat to keep in mind. However, it does not undermine the importance of our results given the prevalence of male leaders in contemporary international politics and throughout world history.

The inherent randomness in large national elections makes regression discontinuity a very promising research approach to use here (Eggers et al., 2015; Bertoli, Dafoe & Trager, 2019). On this point, we left non-democracies out of our analysis to avoid cases where elections may have been rigged. Our reason is that elections can only be considered ‘close’ when they are conducted fairly, and when the outcome cannot be ignored or manipulated. Thus, we limit our analysis to democracies, which we define as countries that had Polity scores of six or higher (Marshall & Gurr, 2018). We also excluded one case where there was credible evidence of large-scale electoral fraud—Kenya 2007 (Kimani, 2008).

We focus specifically on elections where it was clear who the leader would be based on the results. For parliamentary systems, we thus included only those systems where the party that wins the largest number of seats is granted the first opportunity to form the

government and choose the prime minister, either by constitutional procedure or by norms of parliamentary procedure. This meant we excluded parliamentary democracies in which governments are formed by post-election bargaining and coalition-building. The reason for this exclusion is that if the outcome of an election is determined by coalition-building and bargaining, then it is impossible to quantify whether such an election is ‘close’, as the relative number of seats may not determine who forms the government. More information on which parliamentary elections we included is available in the Online appendix (Section A3).

To measure aggression, we use the standard method in political science of looking at the number of militarized interstate disputes that countries initiated. These disputes are cases where countries explicitly threatened, displayed, or used military force against other states (Ghosn, Palmer & Bremer, 2004; Palmer et al., 2022). While many of these disputes are low-level, they are useful for our purposes because they provide an indicator of how hawkish a country’s foreign policy was during a specific period in history. Along with military disputes initiated, we also look at uses of military force, which is included in the Militarized Interstate Dispute dataset (v5.0).

Although the Militarized Interstate Dispute dataset is widely used, some scholars have argued that it contains inaccuracies. Therefore, we checked our data against the revised Militarized Interstate Dispute dataset created by Gibler, Miller & Little (2016). Fortunately, our results were notably stronger for both disputes initiated and uses of force when we used the revised dataset. Nonetheless, we use the standard dataset in the tests in this article to adhere to our pre-analysis plan.

For our main analysis, we estimate the difference at the cut-point and compute p-values using difference-in-means randomization tests, which is the recommended approach for

regression discontinuities with small sample sizes (Cattaneo, Frandsen & Titiunik, 2015; Cattaneo, Titiunik & Vazquez-Bare, 2016). The RD window size (or bandwidth) that we pre-registered was 2%, but we also report the results for other RD window sizes ranging from 1%-5%. In total, there were 34 cases within 2% of the cut-point.

Despite our sample size not being large, the power tests that we ran at the beginning of this project indicated that our results could be informative. In fact, 34 cases constitutes a large enough sample size to detect a medium-sized treatment effect (0.5 x SD) about 31% of the time, a large treatment effect (0.8 x SD) about 59% of the time, and a very large treatment effect (1.2 x SD) about 93% of the time. This is because even though a sample size of 34 is relatively small, it provides enough statistical power to pick up large and very large effects most of the time. However, it is insufficient to detect small and medium size effects reliably. Given the strong association between age and aggression found in other fields, we believed that our design could be sufficiently powered, although we recognized that if we were dealing with a small or medium-sized effect we would likely not have enough data to get conclusive results.

In addition to randomization inference, we also analyze the results using local linear regression. This approach is typically preferred for electoral regression discontinuities (De la Cuesta & Imai, 2016), but it tends to perform poorly when the sample size is small (Cattaneo, Frandsen & Titiunik, 2015; Cattaneo, Titiunik & Vazquez-Bare, 2016). For disputes initiated, our results are very consistent across all of these tests. Our results for uses of force are less stable, particularly when we use local linear regression. This may be because we do not have as many uses of force in our dataset as disputes initiated, and local linear regression often does not perform well when the amount of data is small.

Provided that the outcomes of the close elections in our sample were as-if random, our research design gives us an unbiased estimate of the causal effect of electing older leaders for the countries in our sample. This is what statisticians refer to as internal validity (Imai, 2017). Another important issue to consider is external validity, or how representative the sample is of the broader population of interest (Krcmaric, Nelson & Roberts, 2020). This potential concern is common in experimental research, because most experiments are done on non-random samples from the population. Moreover, even when a researcher does take a random sample from the population, the timing of the experiment, which is chosen by the researcher, could make the sample unrepresentative of past or future cases. Therefore, the question of external validity is a key potential concern for all experiments and natural experiments and not unique to our specific design. However, it is still a very important issue to consider, as we are interested in understanding broad trends in international relations beyond our specific sample.

Fortunately, our sample appears to be fairly representative, at least of other democracies. Figure 1 uses box plots to compare our sample to all other democracies (on the top) and all other countries (on the bottom). Specifically, the comparison in the top panel is between the 34 democracies in our sample (the year of each election) and all 4,087 democracy-years from 1925-2012 in the *Correlates of War* National Material Capabilities data set (v5.0) (Singer, Bremer & Stuckey, 1972). We start with 1925 because Germany (1925) is the earliest case in our sample. As the figure shows, our sample is quite similar to all other democracies. However, it is less representative when we compare it to all other countries, including non-democracies. Democracies and non-democracies also tend to have much different leadership decision-making processes. Therefore, we need to be more cautious about extrapolating from our data to trends for non-democracies. This is true even though

most of the anecdotal evidence that younger leaders tend to be more likely to engage in international conflict actually comes from non-democracies.

Figure 1 in here

We ultimately leave it up to readers to decide for themselves how much our finding for democratic leaders speaks to what might happen in non-democracies. Given that non-democratic leaders are likely to have more power over foreign policy, we do believe that our results should shift priors about how having an older or younger leader affects the likelihood that non-democracies will engage in international conflict.

We also examined the cases where the older and younger leaders barely won to verify that our design achieved comparable treatment and control groups (internal validity). The balance looked very good. In total, we examined 44 pre-treatment variables. The balance between the treatment and control groups looked about as good as we would expect in a randomized experiment. The variables that we looked at included total population ($p \approx 0.91$), urban population ($p \approx 0.76$), iron and steel production ($p \approx 0.75$), military expenditures ($p \approx 0.91$), previous military disputes initiated ($p \approx 0.34$), and previous uses of force ($p \approx 0.37$). These p-values are very reassuring, since in real experiments the p-values for pre-treatment covariates should be distributed uniformly between 0 and 1. We provide a graph illustrating the balance in our data in the Online appendix (Section C2).

Results

The results indicate that electing older leaders tends to make countries much less aggressive than electing younger leaders. Figure 2 provides an illustration of the effect. As countries move from electing younger leaders (on the left) to electing older ones (on the right), they experience a notable change in their propensity to initiate military disputes.

Thus, it appears that electing older leaders makes countries less likely to start military conflicts.

Figure 2 in here

Table I shows the analysis of the data in our pre-registered 2% RD Window. The results are statistically significant for both military disputes initiated ($p \approx 0.005$) and uses of force ($p \approx 0.036$). The table also shows that only 26% of the countries where the older candidate barely won started at least one military dispute, compared to 60% of the countries where the younger candidate barely won ($p \approx 0.080$). Similarly, only 11% of the countries where the older candidate barely won used force at least once, compared to 47% of the countries where the younger candidate barely won ($p \approx 0.025$). This difference is notable because the terms of the older leaders were on average about a year longer than the terms of the younger leaders, so the older leaders had more time to engage in military conflict.

Table I in here

We plot the point estimates and confidence intervals for a number of different RD window sizes in Figures 3 and 4. The results for disputes initiated are statistically significant for RD window sizes of 2% ($p \approx 0.005$) and 3% ($p \approx 0.001$). They are almost statistically significant for an RD window size of 1% ($p \approx 0.058$), despite the statistical power being much lower for this test due to the small sample size. Similarly, the results are close to being statistically significant for RD window sizes of 4% ($p \approx 0.071$) and 5% ($p \approx 0.067$). This lack of significance is not very concerning, though, since these tests include elections that were not plausibly as-if random. Our results for uses of force mirror the overall pattern of disputes initiated. They are statistically significant for RD window sizes of 2% ($p \approx 0.036$) and 3% ($p \approx 0.009$). However, they are not statistically significant for 1%

($p \approx 0.56$), 4% ($p \approx 0.17$), or 5% ($p \approx 0.17$). The lack of significance for these tests should not be too concerning for the reasons already discussed.

Figure 3 in here

Figure 4 in here

Although local linear regression is not reliable for regression discontinuities with small sample sizes, we still applied this approach to the data as an additional test. The results for disputes initiated are statistically significant ($p < 0.015$) for bandwidths from 1%-5%, and the estimated treatment effects are similar. The local linear regression results for uses of force are inconsistent and unstable across the different bandwidths, with nothing being statistically significant. This is likely because local linear regression is an unreliable method in this context given the small number of uses of force in our data.

Our results hold after controlling for a number of economic, military, and demographic factors. We do this as a robustness check, because it can reduce noise caused by small differences between the treatment and control groups that arise by chance. More information about the robustness checks can be found in the Online appendix (Section D).

Recall that one of our major design decisions was to only include cases where the top two candidates were at least ten years apart in age. Figure 5 shows how the estimated treatment effect and confidence intervals change when we set this minimum age difference anywhere between 1 and 20 years. As the figure shows, the size of the estimated effect grows fairly linearly as the age gap between the two candidates increases. This is what we should expect if older candidates tend to be more peaceful than younger ones. Since we would not expect much variation between candidates who were very similar in age, elections with small age differences should simply be adding noise to the dataset. As

we drop these cases, however, we would expect to hone in on the cases where the treatment effect is larger and more visible. Of course, as we drop more and more cases, the confidence intervals should also widen given the decreasing sample size. This appears to be what is happening in the graph.

Figure 5 in here

As a supplementary test, we also looked at whether the estimated treatment effect was larger for presidential democracies that had united governments as opposed to presidential democracies that had divided ones. The reason is that leaders with united governments may have more freedom in foreign policy, which could make it more likely that their personal traits would have a meaningful impact on state behavior (Howell & Pevehouse, 2005, 2007). In fact, we found that the estimated treatment effect was much larger for presidential democracies with united governments. They started about 0.33 fewer disputes per year after electing older leaders, compared to only about 0.14 for presidential democracies with divided governments. This pattern appears for uses of force as well. Thus, the estimated treatment effect is much larger for presidential democracies with united governments.

We found a similar pattern when we compared more consolidated democracies (Polity score = 9 or 10) to less consolidated democracies (Polity score = 6, 7, or 8). The estimated effect of electing older leaders was 0.27 for more consolidated democracies ($p \approx 0.004$, $n=21$), compared to only about 0.07 for less consolidated democracies ($p \approx 0.50$, $n=13$). This difference accords with past research suggesting that leaders' traits and predispositions matter more when democratic leaders have more consolidated political power (Buono de Mesquita et al., 1999; Chiozza & Goemans, 2003; Palmer, London & Regan, 2004).

Lastly, we checked if the older leaders were more or less likely than the younger leaders to be targeted by other countries in military disputes. We found that the two groups were

targeted at about the same rates. Both of these rates were very low. Only two of the older leaders were targeted, along with two of the younger leaders. Thus, having an older leader does not seem to make a country more likely to be attacked by other states.

Considering alternative explanations

Our results show that older leaders are less likely to start military conflicts, controlling for environment. However, it is still important to consider if our findings are simply another example of older people behaving less aggressively than younger ones, or if the story is more complicated here given our setting.

One alternative explanation is that younger leaders may be in office for shorter periods of time. This could lead them to initiate disputes at higher rates simply due to the tendency for leaders to engage in conflict early in their tenures (Potter, 2007; Wolford, 2007; Wu & Wolford, 2018). In our sample, the average term length of younger leaders was about 3.54 years, compared to 4.36 years for older leaders ($p \approx 0.15$). While this is a notable difference, the data suggests that our results are not explained by younger leaders staying in power for shorter periods. First, our findings for both disputes initiated and uses of force remain significant when we control for the amount of time in the leaders term. Second, the results for disputes initiated hold when we look at the total number of disputes that the leaders started in their terms instead of disputes per year ($p \approx 0.024$). It is noteworthy that this test is significant, since the older leaders were on average in power about a year longer than the younger leaders and thus had more time to initiate disputes. In sum, our results do not seem to be driven by younger leaders being in office for shorter periods of time.

What about other differences between the younger and older leaders? Earlier, we discussed a selection process by which politicians rise to the level of potentially becoming

national leaders. This selection process could lead to a distorted pool of younger politicians who are abnormally aggressive and older politicians who are abnormally pacifist. One way to address this possibility is to compare the younger and older leaders on factors other than age and check if there is any indication that there are key differences between the two groups besides age. When we did this, we actually found that the younger and older leaders in our sample were quite similar to each other. They were fairly well-balanced on party ideology ($p \approx 0.86$) and whether they had ever served in the military ($p \approx 0.72$). The younger and older leaders were also well-balanced on the occupational backgrounds that we looked at, including law ($p = 1.00$), journalism ($p \approx 0.57$), and medicine ($p = 1.00$).¹

We did find two notable differences. First, the older leaders in our sample were more likely to have been the incumbent ($p \approx 0.10$). This difference is not too surprising, but it is somewhat concerning because a number of studies have found a link between new leaders and the likelihood and intensity of military conflict (e.g. Smith & Spaniel, 2019). However, we found no evidence that incumbency was driving our results. In fact, both the older new leaders and older incumbents in our data started fewer military disputes per year on average than either the younger new leaders or younger incumbents. After controlling for incumbency, our results remain significant for disputes initiated ($p \approx 0.006$). The findings for uses of force are not quite as robust ($p \approx 0.11$), but the estimated effect remains similar. The second imbalance was that the older leaders were more likely to have combat experience ($p \approx 0.052$). This difference does not appear to explain our findings either. The results for disputes initiated hold after controlling for combat experience ($p \approx 0.011$), and they almost remain significant for uses of force ($p \approx 0.062$). Thus, neither incumbency nor combat experience seem to be playing a pivotal role in our data.

¹ We found the leader characteristic variables primarily in the dataset constructed by Horowitz, Stam & Ellis (2015).

In sum, a closer inspection of the data suggests that our results appear to be another example of older people behaving less aggressively than younger ones, similar to what has been identified in so many other contexts. While the older and younger leaders in our data tended to differ in some ways besides age, none of these differences seem to be driving the results.

Conclusion

In this article, we use a regression discontinuity approach to test whether younger or older leaders tend to make interstate conflict more likely. Consistent with much prior work on age and aggression, we find that older leaders tend to decrease the chances of international conflict. The results are statistically significant for the procedures that we outlined in our pre-analysis plan, and they are also robust to a number of alternative specifications. The effect also seems to be much stronger when presidential democracies have united governments, suggesting that in these cases leader traits may matter more for foreign policy.

Our findings have several important implications for social science research. First, they support the idea that youth may partly explain why so many younger leaders have launched aggressive military campaigns. Our findings should encourage historians and social scientists to take leader age very seriously when investigating the causes of international conflicts. Our results also suggest that the international community should be wary when young politicians come to power. These leaders appear much more willing to engage in military conflict than their older counterparts.

Second, this article highlights the importance of examining how the personal characteristics of leaders can impact international relations. Our large effect-sizes suggest that who rules a country does have a major impact on that country's foreign policy. Therefore,

social scientists should continue searching for leader characteristics that may be important drivers of state behavior. Such studies could shed light on the underlying causes of past wars, as well as make it easier to predict military conflict in the future.

Lastly, this article highlights the value of design-based approaches like regression discontinuity as a complement to more traditional large- N observational analyses. These large- N methods can provide valuable contributions to the study of international relations by identifying robust statistical relationships between important factors. What natural experiments and regression discontinuities bring in is the counterfactual perspective. For example, how might countries' foreign policies differ if they elect older vs. younger leaders? Such design based approaches will likely be most beneficial when researchers are dealing with large effect sizes, because in these cases relatively small- N natural experiments can have adequate statistical power. Such studies should also show that the cases in their samples are fairly representative of the broader population of interest, as we have done in this article.

Replication data: The dataset, codebook, and replication code for the empirical analysis in this article, along with the Online appendix, are available at <https://www.prio.org/jpr/datasets/>. All analyses were conducted using R.

Acknowledgements: We would like to thank the editor at JPR and the anonymous referees for helpful feedback. We would also like to thank Aila Matanock, Alexander Kuo, Ben Olken, Caroline Brandt, Carlos Lastra, Chad Hazlett, Chaim Kaufmann, Daniel Kselman, Daniel Treisman, Daniel Flynn, Evangelos Liaras, Fredrik Sävje, Giacomo Chiozza, Hein Goemans, Henry Pascoe, Irene Menéndez, Kelly Grieco, Margaret Kosal, Michael

Horowitz, Michaela Mattes, Jasjeet Sekhon, Josh Kalla, Michael Becher, Nikitas Konstantinidis, Nina Wiesehomeier, Ron Hassner, and Stephanie Lackner. We would also like to thank everyone who gave us comments at the American Political Science Association Annual Conference, the International Studies Association Annual Conference, the Midwest Political Science Association Annual Conference, the International Relations and Causal Inference Seminars at the University of California, Berkeley, the UC Institute on Great Power Competition in the 21st Century, and the IE University Politics and Economics Research Seminar.

Funding: We gratefully acknowledge support from the University of California Laboratory Fees Grant Program and the Charles Koch Foundation.

References

- Albert, DJ; RH Jonik, Neil V Watson, Boris B Gorzalka & Michael L Walsh (1990) Hormone-dependent aggression in male rats is proportional to serum testosterone concentration but sexual behavior is not. *Physiology & Behavior* 48(3): 409-416.
- Anderson, Craig A; Arlin J Benjamin, Phillip K Wood & Angelica M Bonacci (2006) Development and testing of the Velicer Attitudes toward Violence Scale: Evidence for a four-factor model. *Aggressive Behavior* 52(2): 122-136.
- Barceló, Joan (2020) Are Western-educated leaders less prone to initiate militarized disputes? *British Journal of Political Science* 50(2): 1-32.
- Berman, Mitch; Brian Gladue & Stuart Taylor (1993) The effects of hormones, Type A behavior pattern, and provocation on aggression in men. *Motivation and Emotion* 17(2): 125-138.

- Bertoli, Andrew; Allan Dafoe & Robert F Trager (2019) Is there a war party? Party change, the left-right divide, and international conflict. *Journal of Conflict Resolution* 63(4): 950-975.
- Birditt, Kira S & Karen L Fingerman (2005) Do we get better at picking our battles? Age group differences in descriptions of behavioral reactions to interpersonal tensions. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 60(3): P121-P128.
- Book, Angela S; Katherine B Starzyk & Vernon L Quinsey (2001) The relationship between testosterone and aggression: A meta-analysis. *Aggression and Violent Behavior* 6(6): 579-599.
- Box-Steffensmeier, Janet; Henry E Brady & David Collier (2008) Political science methodology. In: Janet Box-Steffensmeier, Henry E Brady & David Collier (eds) *The Oxford Handbook of Political Methodology* Oxford: Oxford University Press, 3-31.
- Bruine de Bruin, Wändi; Andrew M Parker & Baruch Fischhoff (2012) Explaining adult age differences in decision-making competence. *Journal of Behavioral Decision Making* 25(4): 352-360.
- Bueno De Mesquita, Bruce; James D Morrow, Randolph M Siverson & Alastair Smith (1999) An institutional explanation of the democratic peace. *American Political Science Review* 93(4): 791-807.
- Byman, Daniel L & Kenneth M Pollack (2001) Let us now praise great men: Bringing the statesman back in. *International Security* 25(4): 107-146.
- Carstensen, Laura L (2006) The influence of a sense of time on human development. *Science* 312(5782): 1913-1915.

- Carstensen, Laura L; Helene H Fung & Susan T Charles (2003) Socioemotional selectivity theory and the regulation of emotion in the second half of life. *Motivation and Emotion* 27(2): 103-123.
- Cattaneo, Matias D; Brigham R Frandsen & Rocio Titiunik (2015) Randomization inference in the regression discontinuity design: An application to party advantages in the US Senate. *Journal of Causal Inference* 3(1): 1-24.
- Cattaneo, Matias D; Rocio Titiunik & Gonzalo Vazquez-Bare (2016) Inference in regression discontinuity designs under local randomization. *Stata Journal* 16(2): 331-367.
- Chiozza, Giacomo & Hein E Goemans (2003) Peace through insecurity: Tenure and international conflict. *Journal of Conflict Resolution* 47(4): 443-467.
- Connor, Walter Robert (1984) *Thucydides*. Princeton, NJ: Princeton University Press.
- Daly, Martin & Margo Wilson (1990) Killing the competition. *Human Nature* 1(1): 81-107.
- De la Cuesta, Brandon & Kosuke Imai (2016) Misunderstandings about the regression discontinuity design in the study of close elections. *Annual Review of Political Science* 19: 375-396.
- Dube, Oeindrila & SP Harish (2020) Queens: A study of gender and war in historical Europe. *Journal of Political Economy* 128(7): 2579-2652.
- Edwards, David A (1969) Early androgen stimulation and aggressive behavior in male and female mice. *Physiology & Behavior* 4(3): 333-338.
- Eggers, Andrew; Anthony Fowler, Jens Hainmueller, Andrew B Hall & James M Snyder Jr (2015) On the validity of the regression discontinuity design for estimating electoral effects: New evidence from over 40,000 close races. *American Journal of Political Science* 59(1): 259-274.

- Eisenegger, Cristoph; Michael Naef, Philippe D Snozzi, Markus Heinrichs & Ernst Fehr (2010) Prejudice and truth about the effect of testosterone on human bargaining behaviour. *Nature* 463(7279): 356-359
- Fabbri, Elisa; Yang An, Marta Gonzalez-Freire, Marco Zoli, Marcello Maggio, Stephanie A Studenski, Josephine M Egan, Chee W Chia & Luigi Ferrucci (2016) Bioavailable testosterone linearly declines over a wide age spectrum in men and women from the Baltimore Longitudinal Study of Aging. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences* 71(9): 1202-1209.
- Farrington, David P (1986) Age and crime. *Crime and Justice* 7: 189-250.
- Fischer, Hakan; Johan Sandblom, Joachim Gavazzeni, Peter Fransson, Christopher Wright & Lars Backman (2005) Age-differential patterns of brain activation during perception of angry faces. *Neuroscience Letters* 386(2): 99-104.
- Ghosn, Faten; Glenn Palmer & Stuart A Bremer (2004) The MID3 data set, 1993-2001: Procedures, coding rules, and description. *Conflict Management and Peace Science* 21(2): 133-54.
- Gibler, Douglas M; Steven V Miller & Erin K Little (2016) An analysis of the Militarized Interstate Dispute (MID) dataset, 1816-2001. *International Studies Quarterly* 60(4): 719-730.
- Hay, Elizabeth L & Manfred Diehl (2011) Emotion complexity and emotion regulation across adulthood. *European Journal of Aging* 8(3): 157-168.
- Horowitz, Michael C; Rose McDermott & Allan C Stam (2005) Leader age, regime type, and violent international relations. *Journal of Conflict Resolution* 49(5): 661-685.

- Horowitz, Michael C; Philip Potter, Todd S Sechser & Allan Stam (2018) Sizing up the adversary: Leader attributes and coercion in international conflict. *Journal of Conflict Resolution* 62(10): 2180-2204.
- Horowitz, Michael C; Allan C Stam & Cali M Ellis (2015) *Why Leaders Fight*. New York: Cambridge University Press.
- Howell, William G & Jon C Pevehouse (2005) Presidents, Congress, and the use of force. *International Organization* 59(1): 209-232.
- Howell, William G & Jon C Pevehouse (2007) *While Dangers Gather: Congressional Checks on Presidential War Powers*. Princeton, NJ: Princeton University Press.
- Imai, Kosuke (2017) *Quantitative Social Science: An Introduction*. Princeton, NJ: Princeton University Press.
- Isaacowitz, Derek M; Kaitlin Toner, Deborah Goren & Hugh R Wilson (2008) Looking while unhappy: Mood-congruent gaze in young adults, positive gaze in older adults. *Psychological Science* 19(9): 848-853.
- Jenkins, Roy (2001) *Churchill*. New York: Penguin Books.
- Jones, Benjamin F & Benjamin A Olken (2009) Hit or miss? The effect of assassinations on institutions and war. *American Economic Journal: Macroeconomics* 1(2): 55-87.
- Kalmoe, Nathan P (2013) From fistfights to firefights: Trait aggression and support for state violence. *Political Behavior* 35: 311-30.
- Kimani, Mary (2008) East Africa feels blows of Kenyan crisis. *United Nations* April (<https://www.un.org/africarenewal/magazine/april-2008/east-africa-feels-blows-kenyan-crisis>).
- Kisley, Michael A; Stacey Wood & Christina L Burrows (2007) Looking at the sunny side of life: Age-related change in an event-related potential measure of the negativity bias. *Psychological Science* 18(9): 838-843.

- Knight, Marisa; Travis L Seymour, Joshua T Gaunt, Christopher Baker, Kathryn Nesmith & Mara Mather (2007) Aging and goal-directed emotional attention: Distraction reverses emotional biases. *Emotion* 7(4): 705-714.
- Krcmaric, Daniel; Stephen C Nelson & Andrew Roberts (2020) Studying leaders and elites: The personal biography approach. *Annual Review of Political Science* 23: 133-151.
- Leclerc, Christina M & Elizabeth A Kensinger (2011) Neural processing of emotional pictures and words: A comparison of young and old adults. *Developmental Neuropsychology* 36(4): 519-538.
- MacLeod, John F; Peter G Grove & David P Farrington (2012) *Explaining Criminal Careers: Implications for Justice Policy*. Oxford: Oxford University Press.
- Marquez-González, Mária; Mária Izal Fernández de Trocóniz, Ignacio Montorio Cerrato & Andrés Losada Baltar (2008) Emotional experience and regulation across the adult lifespan: Comparative analysis in three age groups. *Psicothema* 20(4): 616-622.
- Marshall, Monty G & Ted Robert Gurr (2018) Political regime characteristics and transitions, 1800-2018. Center for Systemic Peace (<http://www.systemicpeace.org/inscrdata.html>).
- Mather, Mara (2012) The emotion paradox in the aging brain. *Annals of the New York Academy of Sciences* 1251(1): 33-49.
- Mather, Mara; Turhan Canli, Tammy English, Sue Whitfield, Peter Wais, Kevin Ochsner, DE Gabrieli John & Laura Carstensen (2004) Amygdala responses to emotionally valenced stimuli in older and younger adults. *Psychological Science* 15(4): 259-263.

- McDermott, Rose & Peter K Hatemi (2017) The relationship between physical aggression, foreign policy and moral choices: Phenotypic and genetic findings. *Aggression and Behavior* 43: 37-46.
- McIntyre, Matthew H; Emily S Barrett, Rose McDermott, Dominic DP Johnson, Jonathan Cowden & Stephen P Rosen (2007) Finger length ratio (2D: 4D) and sex differences in aggression during a simulated war game. *Personality and Individual Differences* 42(4): 755-764.
- Mearsheimer, John J (2014) America unhinged. *The National Interest* 129: 9-30.
- Mearsheimer, John J & Stephen M Walt (2003) An unnecessary war. *Foreign Policy* 134: 50-9.
- Neupert, Shevaun D; David M Almeida & Susan Turk Charles (2007) Age differences in reactivity to daily stressors: The role of personal control. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 62(4): P216-P225.
- Nyseth Brehm, Hollie; Christopher Uggen & Gasanabo Jean-Damascène (2016) Age, gender, and the crime of crimes: Toward a life-course theory of genocide participation. *Criminology* 54(4): 713-43.
- Palmer, Glenn; Tamar London & Patrick Regan (2004) What's stopping you? The sources of political constraints on international conflict behavior in parliamentary democracies. *International Interactions* 30(1): 1-24.
- Palmer, Glenn; Roseanne W McManus, Vito D'Orazio, Michael R Kenwick, Mikaela Karstens, Chase Bloch, Nick Dietrich, Kayla Kahn, Kellan Ritter & Michael J Soules (2022) The MID5 dataset, 2011-2014: Procedures, coding rules, and description. *Conflict Management and Peace Science* 39(4): 470-482.

- Potter, Philip BK (2007) Does experience matter? American Presidential experience, age, and international conflict. *Journal of Conflict Resolution* 51(3): 351-378.
- Rich, Norman (1985) *Why the Crimean War? A Cautionary Tale*. Hanover, NH: University Press for New England.
- Robinson, Gregory; John E McNulty & Jonathan S Krasno (2009) Observing the counterfactual? The search for political experiments in nature. *Political Analysis* 17(4): 341-357.
- Saunders, Elizabeth N (2011) *Leaders at War: How Presidents Shape Military Interventions*. Ithaca, NY: Cornell University Press.
- Siennick, Sonja E & D Wayne Osgood (2008) A review of research on the impact of crime on transitions to adult roles. In: Akiva M Liberman (ed.) *The Long View of Crime: A Synthesis of Longitudinal Research*. New York: Springer, 161-187.
- Siever, Larry J (2008) Neurobiology of aggression and violence. *American Journal of Psychiatry* 165(4): 429-442.
- Singer, J David; Stuart Bremer & John Stuckey (1972) Capability distribution, uncertainty, and major power war, 1820-1965. In: Bruce Russett (ed.) *Peace, War, and Numbers*. Beverly Hills, CA: Sage, 19-48.
- Smith, Bradley C & William Spaniel (2019) Militarized disputes, uncertainty, and leader tenure. *Journal of Conflict Resolution* 63(5): 1222-1252.
- Stawski, Robert S; Jacqueline Mogle & Martin J Sliwinski (2011) Intraindividual coupling of daily stressors and cognitive interference in old age. *Journal of Gerontology Series B: Psychological Sciences and Social Sciences* 66: 121-129.
- Steffensmeier, Darrell & Cathy Streifel (1991) Age, gender, and crime across three historical groups. *Social Forces* 69(3): 869-894.

- Steffensmeier, Darrell J; Emilie Andersen Allan, Miles D Harer & Cathy Streifel (1989) Age and the distribution of crime. *American Journal of Sociology* 94(4): 803-31.
- Sweeten, Gary; Alex R Piquero & Laurence Steinberg (2013) Age and the explanation of crime, revisited. *Journal of Youth and Adolescence* 42(6): 921-938.
- Touron, Dayna R & Christopher Hertzog (2004) Distinguishing age differences in knowledge, strategy use, and confidence during strategic skill acquisition. *Psychology and Aging* 19(3): 452-466.
- Trager, Robert F (2012) Long-term consequences of aggressive diplomacy: European relations after Austrian Crimean war threats. *Security Studies* 21(2): 232-265.
- Van Honk, Jack; Estrella R Montoya, Peter A Bos, Mark van Vugt & David Terburg (2012) New evidence on testosterone and cooperation. *Nature* 485(7399): 356-359.
- Waltz, Kenneth N (1979) *Theory of International Politics*. Long Grove, IL: Waveland Press.
- Weeks, Jessica (2014) *Dictators at War and Peace*. Ithaca, NY: Cornell University Press.
- Williams, Leanne M; Kerri J Brown, Donna Palmer, Belinda J Liddell, Andrew H Kemp, Gloria Olivieri, Anthony Peduto & Evian Gordon (2006) The mellow years? Neural basis of improving emotional stability over age. *The Journal of Neuroscience* 26(24): 6422-6430.
- Wolford, Scott (2007) The turnover trap: New leaders, reputation, and international conflict. *American Journal of Political Science* 51(4): 772-788.
- Wu, Cathy Xuanxuan & Scott Wolford (2018) Leaders, states, and reputations. *Journal of Conflict Resolution* 62(10): 2087-2117.

ANDREW BERTOLI, PhD in Political Science (UC Berkeley, 2016); Assistant Professor, IE University (2018-).

ALLAN DAFOE, PhD in Political Science (UC Berkeley, 2012); President of the Centre for the Governance of AI (2021-).

ROBERT TRAGER, PhD in Political Science (Columbia University, 2007); Professor, UCLA (2022-).

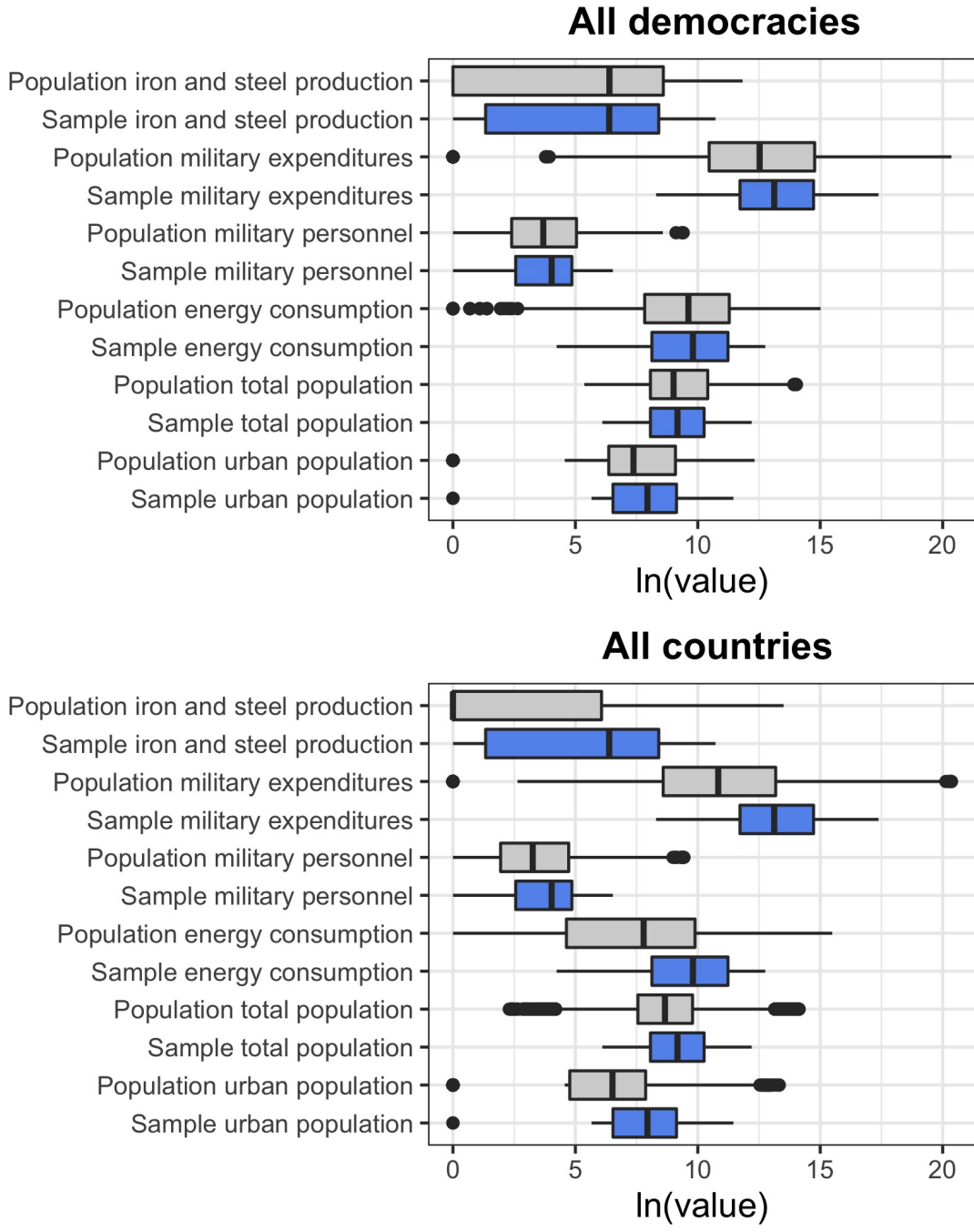


Figure 1. Checking for external validity

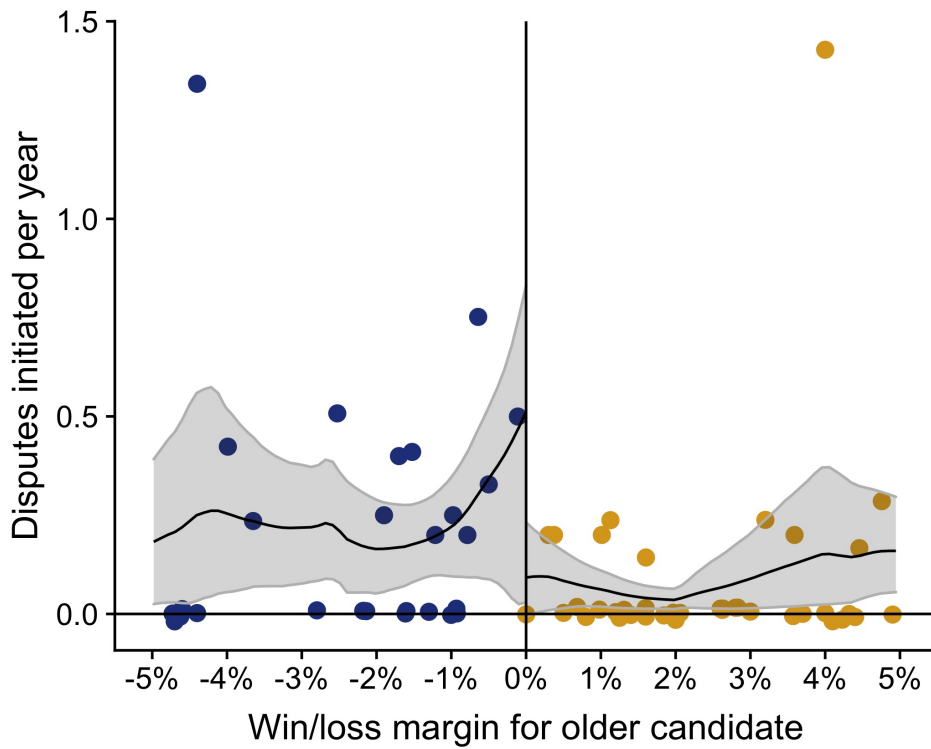


Figure 2. Regression discontinuity graph

The lines in this figure were constructed using local linear regression with a 2% bandwidth. We also jittered the points in each graph to make it easier to see the points on the x-axis (the countries that did not initiate any military disputes). The confidence intervals were constructed using non-parametric bootstrapping.

Table I. Results for the cases in the 2% RD window

Winner	Avg age	Avg number of disputes initiated per year	Percentage of countries that initiated at least one dispute	Avg number of uses of force per year	Percentage of countries that used force at least once
Older leader (n=19)	67	0.05**	26% [†]	0.03*	11%*
Younger leader (n=15)	49	0.22**	60% [†]	0.13*	47%*
p-value		0.005	0.080	0.036	0.025

[†] $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. The p-values in this table come from two-tailed randomization inference tests. A full breakdown of which countries in the sample initiated military disputes and used force against other countries is available in the Online appendix (Section B).

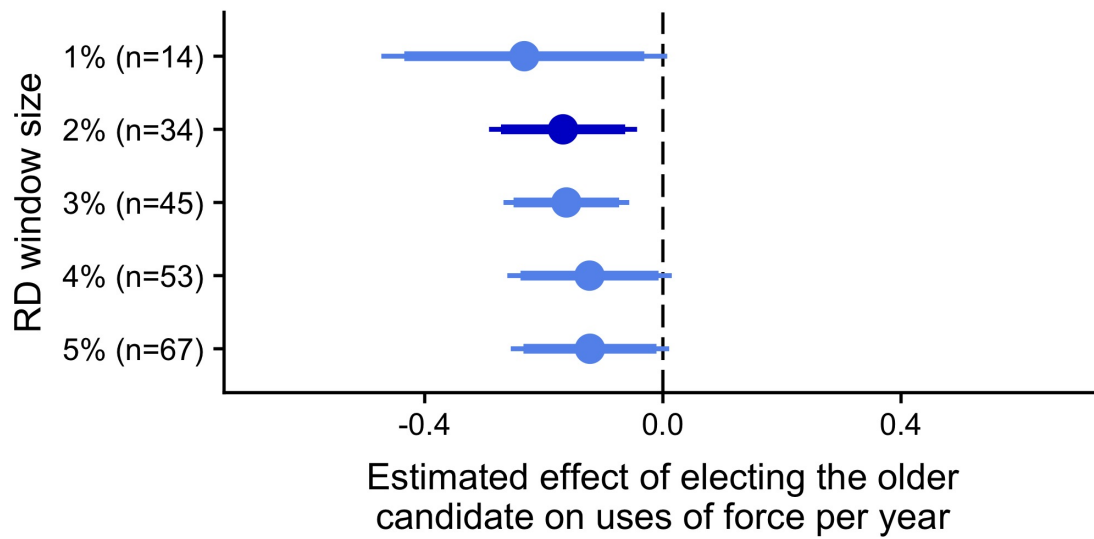


Figure 3. Changing the RD window size (disputes initiated)

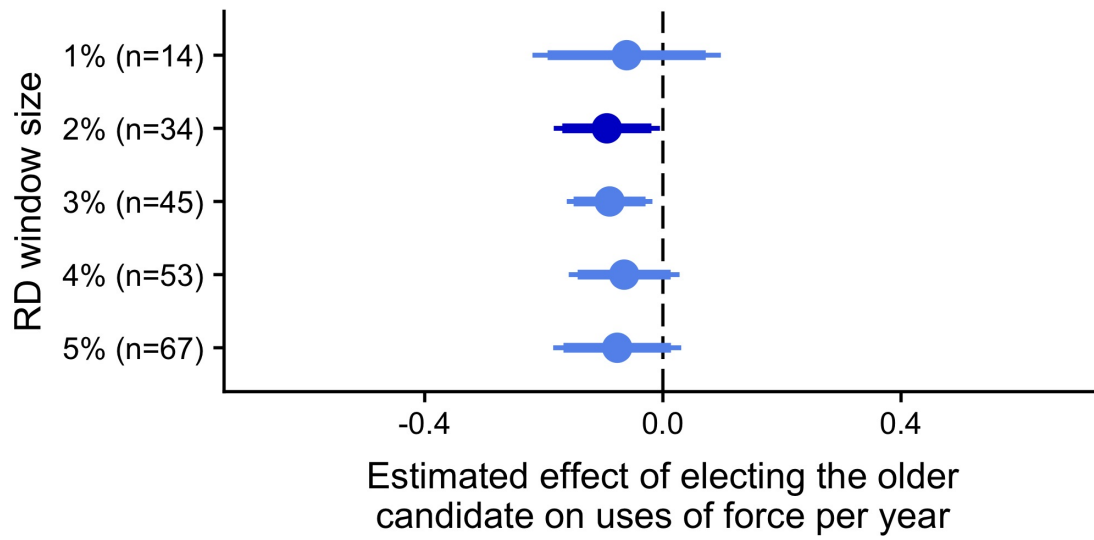


Figure 4. Changing the RD window size (uses of force)

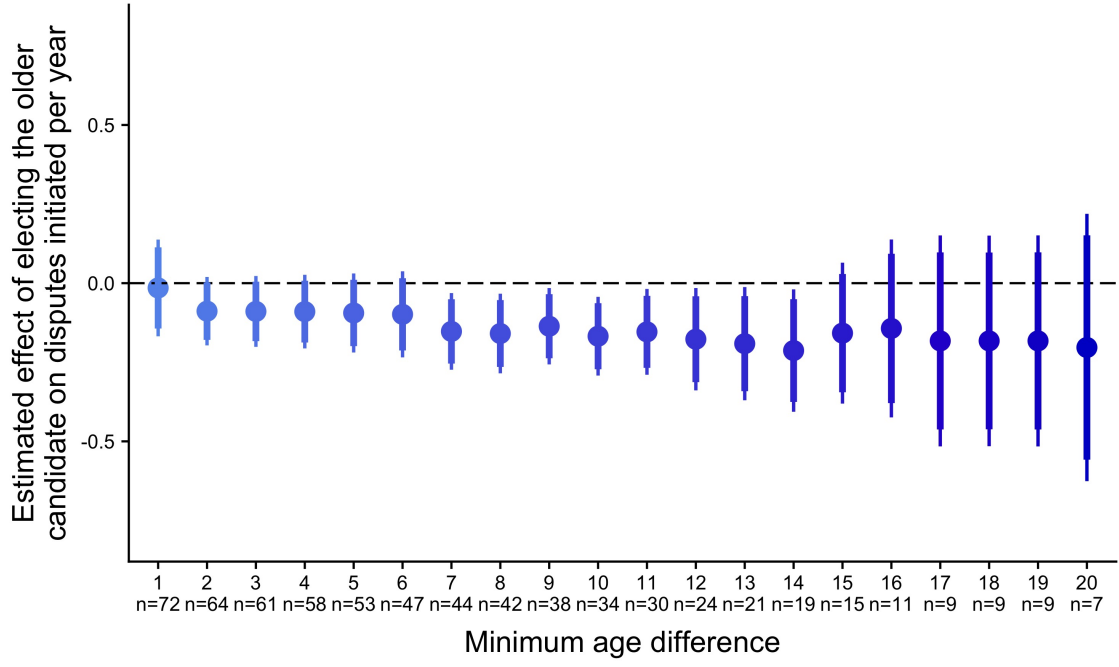


Figure 5. Shifting the minimum age difference