Nationalism and Interstate Conflict: Supporting Information

January 27, 2017

Table of Contents

1 More About Sports, Nationalism, and Conflict 2

1.1 Orwell and Hitchens on International Sports . . . 2
1.2 Is Sports Nationalism Really Exogenous? . . . 4

2 Further Information About the World Cup Results 5

2.1 Additional Balance Plots . . . 5
2.2 Linear Regression . . . 9
2.3 Adjusting the Size of the Regression Discontinuity Window . . . 11
2.4 Shifting the Minimum Score Requirement . . . 12
2.5 Changing the Time Interval . . . 13
2.6 Changing the Cut-point . . . 15
2.7 Addressing SUTVA Violations . . . 16
2.8 Performance of the Qualifiers at the World Cup . . . 18
2.9 Putting the Effect Size in Perspective . . . 19

3 Regional Championships Analysis 20

3.1 Constructing the Sample . . . 20
3.2 Checking for Balance . . . 22
3.3 Additional Information About the Data . . . 26
3.4 Tracking Aggression Levels . . . 27
3.5 Continuity Graph . . . 28
3.6 Linear Regression . . . 29
3.7 Adjusting the Size of the Regression Discontinuity Window . . . 31
3.8 Shifting the Minimum Score Requirement . . . 32
3.9 Changing the Cut-point . . . 33

4 Combined Sample 34

4.1 Balance Plot . . . 34
4.2 Continuity Graph . . . 36

5 Information About the Covariates 37

6 References 38

More About Sports, Nationalism, and Conflict

1.1 Orwell and Hitchens on International Sports

I mention in the paper that George Orwell and Christopher Hitchens wrote about the dangers of sports nationalism, but I lacked the space to discuss their essays in much detail. I provide more information about their articles below.

Orwell published his essay in 1945, titled "The Sporting Spirit". It was a direct response to the Moscow Dynamo soccer trip to Britain following World War II, which I discuss in the paper. A popular theory in Britain at the time was that international sporting events encouraged peace between countries, but these games had the exact opposite effect. In fact, British officials cancelled them when they realized that they were hurting Anglo-Soviet relations. After rehashing the damage that international sports caused in this particular case, Orwell launches into a tirade about the negative impact of international sporting events on world politics: "I am always amazed when I hear people saying that sport creates goodwill between the nations... concrete examples (the 1936 Olympic Games, for instance) [show] that international sporting contests lead to orgies of hatred." He continues,

> At the international level sport is frankly mimic warfare [played by countries] who work themselves into furies over these absurd contests, and seriously believe - at any rate for short periods - that running, jumping and kicking a ball are tests of national virtue... If you wanted to add to the vast fund of ill-will existing in the world at this moment, you could hardly do it better than by a series of football matches between Jews and Arabs, Germans and Czechs, Indians and British, Russians and Poles, and Italians and Jugoslavs.

Moreover, Orwell believed that nationalism is the key factor that makes international sporting events so divisive. As he explains, "There cannot be much doubt that

the whole thing is bound up with the rise of nationalism – that is, with the lunatic modern habit of identifying oneself with large power units and seeing everything in terms of competitive prestige."

Hitchens published his article, "Why the Olympics and Other Sports Cause Conflict", in 2010. His views mirror those of Orwell: "Whether it's the exacerbation of national rivalries that you want–as in Africa this year–or the exhibition of the most depressing traits of the human personality... you need only look to the wide world of sports." He goes on,

> Putting it a bit strongly, you say. But what about the border war between El Salvador and Honduras in 1969, when the violence set off by a disputed soccer match escalated to the point of aerial bombardment? In Khartoum recently, a soccer game between Egypt and Algeria led to widespread violence, a sharp exchange of diplomatic notes, a speech about affronted national honor from President Hosni Mubarak, hysterical hatred pumped out on state media, and an all-round deterioration of what you might call civility. And this between two members of the Arab League!

Along with inciting international conflict, Hitchens argues that sports pose other threats to society. He contends that they distract people from politics, divert money from important public projects, and dumb down the national discourse (due to the widespread use of sports metaphors). He also claims that they can worsen ethnic or regional rivalries within countries when played at the subnational level. Developing and testing these ideas in a rigorous way could make for some interesting future research projects.

1.2 Is Sports Nationalism Really Exogenous?

In the paper, I discuss how World Cup nationalism affects some countries more than others based on how seriously their populations take soccer. I also point out that certain leaders are particularly opportunistic when it comes to encouraging and exploiting sports nationalism. The fact that domestic factors influence the amount of nationalism that international sports create raises an important question: Is sports nationalism really exogenous?

The answer is straightforward: international sports provide countries with exogenous surges of nationalism, the intensity of which they endogenously determine. This issue may sound like a problem, but in fact, virtually all real experiments have units that vary in terms of the degree to which they respond to the treatment. For example, some subjects in a medical experiment might have a gene that blocks the treatment from being absorbed, and others might have a gene that accentuates the treatment. These unit-level differences do not prevent researchers from determining whether the treatment affected the sample.

The same holds true here. Some countries may be largely immune to the effects of sports nationalism, while others may be hypersensitive. Nevertheless, we can still test the average effect of World Cup participation on state aggression for the countries in our sample. In fact, the unit-level variation can actually be useful from a research design standpoint. In the paper, I used it to provide evidence that the design worked by showing that the treatment effect was entirely driven by countries where soccer is the most popular sport.

Further Information About the World Cup Results

2.1 Additional Balance Plots

Figure 1a shows an expanded balance plot that includes some variables not discussed in the paper. The p-values still appear to be distributed uniformly between 0 and 1, which is what would be expected if qualification was random for the countries close to the cut-point. The one potential concern is that the differences in means for military expenditures and military personnel are large. However, the p-values indicate that the qualifier and non-qualifier groups are not imbalanced on these factors. The large differences in means result from the fact that the United States and Soviet Union are outliers and both appear in the qualifier group. This issue alone does not point to a failure of the design. It would arise 25% of the time by chance if barely qualifying or barely missing the World Cup was perfectly random. The results also remain significant when the United States and Soviet Union are dropped from the sample (p=0.021), which alleviates concerns that a difference in military factors is driving the results. In fact, when the United States and Soviet Union are dropped, the non-qualifiers have higher averages for the military factors. Balance for the sample with the United States and Soviet Union removed is presented in Figure 2a.

The qualifiers and non-qualifiers are also well-balanced when the regression discontinuity window is set at 1 point, as Figure 3a shows. The non-qualifiers again have higher averages for the military factors. As mentioned in the paper, the results remain significant for this new sample (p=0.040, n=92).

Figure 1a. Balance Between the Qualifiers and Non-Qualifiers

Variable Name	Treatment Mean	Control Mean
Iron and Steel Production	9,449,340	6,111,800
Military Expenditures	10,809,200,000	3,559,410,000
Military Personnel	310,873	219,817
Total Population	33,438,500	35,742,500
Urban Population	9,189,200	8,879,540
Birth Rate	23.6	24
Death Rate	11.1	11.1
Infant Mortality	51.1	50.6
Energy Production	182,313,000	107,826,000
Imports	23,614,700,000	18,628,100,000
Exports	21,695,600,000	18,894,900,000
Land Area	460,716	188,522
Material Power Score	0.01	0.009
Level of Democracy	0.465	0.465
Great Power Status	0.113	0.113
Engaged in Civil War	0.01	0.01
Resolved Civil War	0.01	0
Year of State Formation	1881	1877
Sex Ratio	96.7	98.8
Life Expectancy	65.6	66
Median Age	27.7	27.7
Number of Alliances	15.7	16.5
U.S. Ally	0.4	0.479
Soccer Most Popular Sport	0.93	0.93
Appearance at Previous World Cup	0.352	0.31
MIDs Initiated in the Year Before	0.183	0.113
MIDs Initiated in the 3 Years Before	0.69	0.577
MIDs Initiated in the 5 Years Before	1.07	0.873

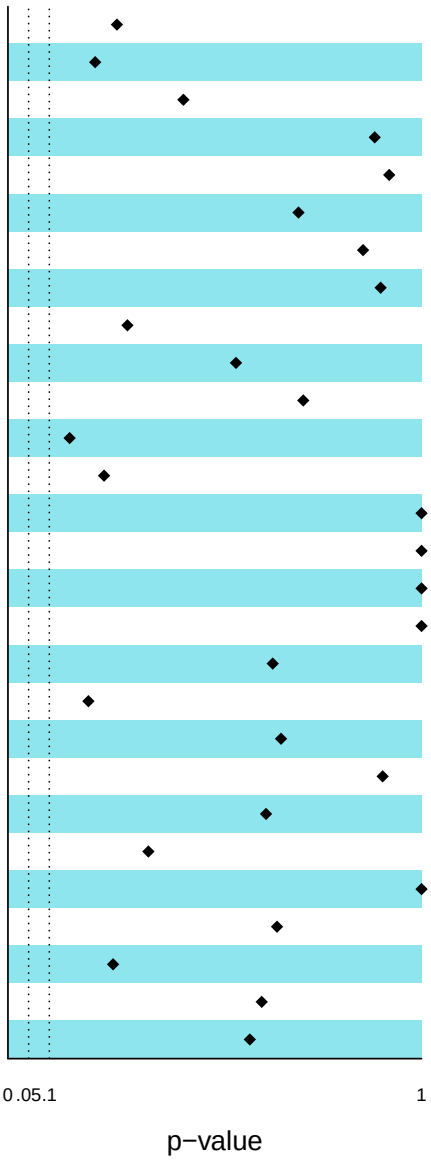


Figure 2a. Balance After Dropping the U.S. and Soviet Union

Variable Name	Treatment Mean	Control Mean
Iron and Steel Production	5,544,190	6,111,800
Military Expenditures	2,497,500,000	3,559,410,000
Military Personnel	168,118	219,817
Total Population	24,210,200	35,742,500
Urban Population	6,348,310	8,879,540
Birth Rate	23.8	24
Death Rate	11.2	11.1
Infant Mortality	51.7	50.6
Energy Production	87,843,000	107,826,000
Imports	16,547,900,000	18,628,100,000
Exports	16,586,500,000	18,894,900,000
Land Area	170,848	188,522
Material Power Score	0.007	0.009
Level of Democracy	0.471	0.465
Great Power Status	0.07	0.113
Engaged in Civil War	0.01	0.01
Resolved Civil War	0.01	0
Year of State Formation	1882	1877
Sex Ratio	97.1	98.8
Life Expectancy	65.5	66
Median Age	27.7	27.7
Number of Alliances	15.6	16.5
U.S. Ally	0.412	0.479
Soccer Most Popular Sport	0.941	0.93
Appearance at Previous World Cup	0.353	0.31
MIDs Initiated in the Year Before	0.162	0.113
MIDs Initiated in the 3 Years Before	0.456	0.577
MIDs Initiated in the 5 Years Before	0.691	0.873

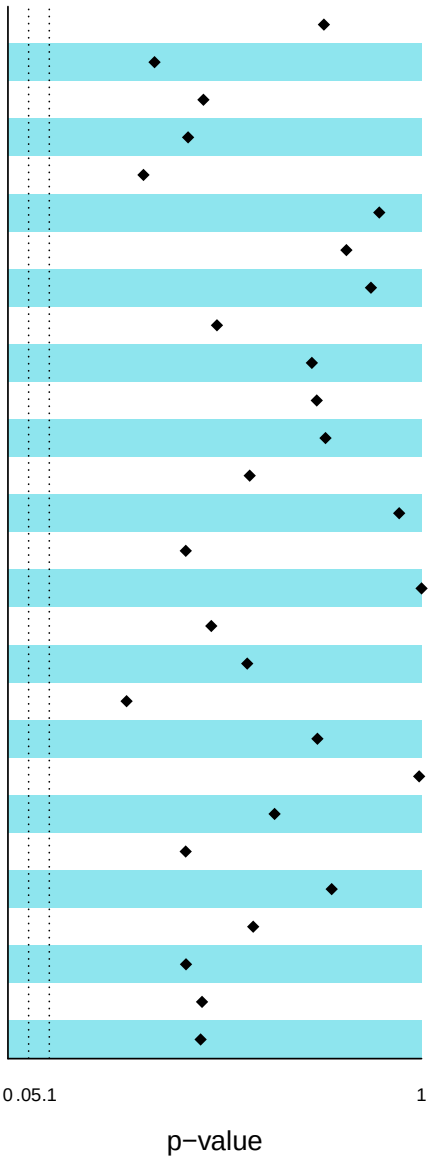
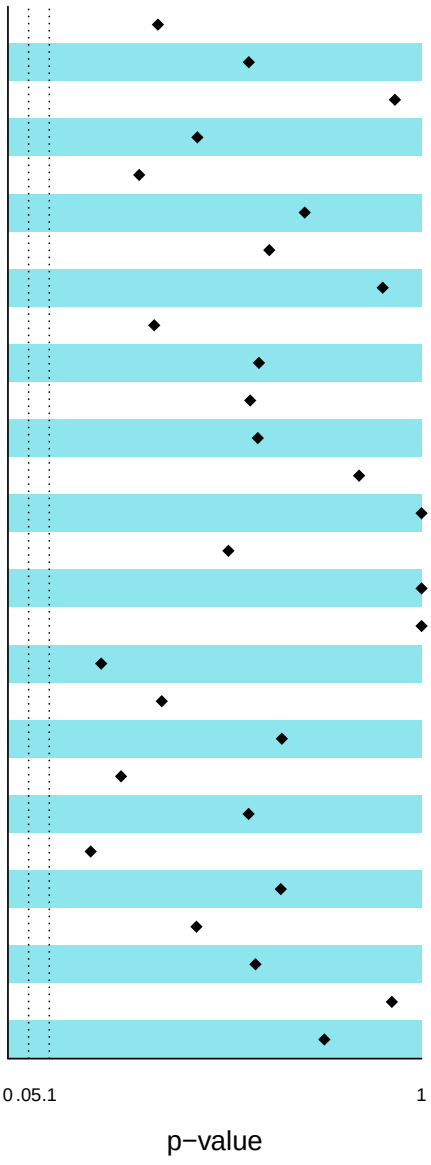


Figure 3a. Balance for the One-Point Window

Variable Name	Treatment Mean	Control Mean
Iron and Steel Production	5,427,090	7,439,200
Military Expenditures	3,299,140,000	4,178,710,000
Military Personnel	247,913	259,478
Total Population	30,092,500	47,563,500
Urban Population	7,477,110	11,235,200
Birth Rate	25	24.5
Death Rate	11.2	11.4
Infant Mortality	54.6	54.1
Energy Production	98,028,100	132,946,000
Imports	17,610,400,000	21,092,700,000
Exports	18,008,600,000	22,709,000,000
Land Area	346,137	238,980
Material Power Score	0.009	0.01
Level of Democracy	0.391	0.391
Great Power Status	0.09	0.13
Engaged in Civil War	0.02	0.02
Resolved Civil War	0.02	0
Year of State Formation	1886	1875
Sex Ratio	97.5	99.7
Life Expectancy	64.8	65.2
Median Age	26.6	27.4
Number of Alliances	15.8	16.9
U.S. Ally	0.326	0.435
Soccer Most Popular Sport	0.935	0.957
Appearance at Previous World Cup	0.261	0.348
MIDs Initiated in the Year Before	0.174	0.13
MIDs Initiated in the 3 Years Before	0.543	0.565
MIDs Initiated in the 5 Years Before	0.761	0.87



2.2 Linear Regression

Table 1a shows the estimated treatment effect after controlling for baseline differences between the two groups, which is a commonly used robustness check for natural experiments. The highest p-value in any of these models is (p=0.023), so the results do not come close to falling out of statistical significance. Also, note that the estimated treatment effect does not change much after controlling for the baseline characteristics. This consistency is expected in natural experiments where as-if randomization worked, since treatment assignment should be independent of the covariates.

While Table 1a uses the difference-in-differences estimator, Table 2a shows the estimated treatment effect controlling for the previous outcome. Whenever the previous outcome is available, researchers can choose to run the statistical tests using the change (difference-in-differences) or just controlling for the previous outcome. Using the change does control for the previous outcome, but it fixes the coefficient at one.

Difference-in-Differences Model

$$\text{Change} = \beta_0 + \beta_1 \cdot \text{Treat} + ... + \varepsilon$$
$$\text{Outcome - Previous Outcome} = \beta_0 + \beta_1 \cdot \text{Treat} + ... + \varepsilon$$
$$\text{Outcome} = \beta_0 + \beta_1 \cdot \text{Treat} + 1 \cdot \text{Previous Outcome} + ... + \varepsilon$$

Controlling for the Previous Outcome

$$\text{Outcome} = \beta_0 + \beta_1 \cdot \text{Treat} + \beta_2 \cdot \text{Previous Outcome} + ... + \varepsilon$$

As these equations show, the two approaches are equivalent except for how they treat the coefficient on the previous outcome. Difference-in-differences fixes it at one, while controlling for the previous outcome uses least squares to estimate the value of β_2. Statisticians disagree over which of these approaches is best, so I provide the results for both models here.

Table 1a. Estimating the Effect of the World Cup on Dispute Initiation with Difference-in-Differences OLS Regression

	Model 1	Model 2	Model 3	Model 4
World Cup Appearance	0.38**	0.29*	0.37**	0.40**
	(0.14)	(0.13)	(0.13)	(0.14)
Military Controls		X	X	X
Economic Controls			X	X
Demographic Controls				X

Table 2a. Estimating the Effect of the World Cup on Dispute Initiation Controlling for the Previous Outcome

	Model 1	Model 2	Model 3	Model 4
World Cup Appearance	0.40**	0.28*	0.31**	0.32**
	(0.13)	(0.11)	(0.11)	(0.11)
Military Controls		X	X	X
Economic Controls			X	X
Demographic Controls				X

Table 3a. Adjusting the Regression Discontinuity Window

Max Made/Missed	Estimated Effect	p-value	n
0 points	0.25	0.503	40
1 point	0.37*	0.027	92
2 points	0.38**	0.007	142
3 points	0.49**	0.001	164
4 points	0.35*	0.011	206
5 points	0.34*	0.012	214
6 points	0.32*	0.012	224
7 points	0.32*	0.012	224
8 points	0.32*	0.012	224
9 points	0.32*	0.012	226

2.3 Adjusting the Size of the Regression Discontinuity Window

As discussed in the paper, the findings are also robust to shrinking or expanding the size of the regression discontinuity window. In fact, the results remain significant if the window is set at one point or anywhere above.

Table 4a. Shifting the Minimum Score Requirement

Minimum Score	Estimated Effect	p-value	n
1 point	0.37*	0.013	156
2 points	0.38*	0.014	154
3 points	0.49*	0.013	148
4 points	0.35*	0.016	146
5 points	0.34**	0.007	142
6 points	0.32*	0.013	126
7 points	0.32*	0.049	90
8 points	0.32	0.083	76
9 points	0.32	0.082	70

2.4 Shifting the Minimum Score Requirement

Recall that when constructing the sample, I only included dyads where the qualifying team scored at least five points in the standings. As Table 4a shows, this minimum score requirement can be set anywhere between zero and seven, and the results remain significant at the 5% level. When the minimum score requirement is set at eight, the sample size drops to 76, and the results are barely insignificant at the 5% level. However, this loss of significance should not raise concern given the decreasing sample size.

2.5 Changing the Time Interval

For the statistical tests in the paper, I use the change in aggression between (1) the period ranging from qualification to the second year after the World Cup and (2) the period of the same length prior to qualification. In other words, I use the change in aggression between the two-and-a-half years before and after qualification. Using this time period before qualification provides a stable baseline measure of aggression for each country in the sample, while using this time period after accounts for residual nationalism and reoccurring disputes. However, the results are largely insensitive to these choices of interval length. As Table 5a shows, the results are significant for almost any choice of interval length prior to qualification, and they remain significant for any choice of interval length following qualification except 1 and 1.5 years.

Moreover, as Table 6a shows, the results are significant for all choices of interval length if the outcome is changed to revisionist disputes initiated. This measure does not include cases where states used force to preserve the status quo, so it is a less noisy indicator of aggression than all disputes initiated (Gowa 1998, 313). Thus, while it is potentially concerning that the results are not significant for all choices of interval length between 0.5 and 3 years, I believe that this is due to noise in the standard measure of aggression rather than the absence of a real treatment effect.

Table 5a. Changing the Length of the Time Interval (Initiated MIDs)

Period Before Qualification			Period After Qualification		
Years	Estimate	p-value	Years	Estimate	p-value
0.5	0.44*	0.010	0.5	0.44	0.054
1.0	0.26	0.148	1.0	0.30	0.134
1.5	0.32*	0.022	1.5	0.32	0.108
2.0	0.35*	0.012	2.0	0.33*	0.033
2.5	0.38**	0.007	2.5	0.38**	0.007
3.0	0.34*	0.010	3.0	0.38*	0.013

Table 6a. Changing the Length of the Time Interval (Revisionist MIDs)

Period Before Qualification			Period After Qualification		
Years	Estimate	p-value	Years	Estimate	p-value
0.5	0.37*	0.045	0.5	0.51**	0.003
1.0	0.30*	0.044	1.0	0.33*	0.029
1.5	0.34**	0.003	1.5	0.37*	0.034
2.0	0.37**	0.001	2.0	0.33**	0.009
2.5	0.38**	0.001	2.5	0.38**	0.001
3.0	0.30*	0.010	3.0	0.33**	0.003

Table 7a. Shifting the Location of the Cut-point

Location of Cut-point	Estimated Effect	p-value	n
-5.5	-0.08	1.000	30
-4.5	-0.47	0.335	41
-3.5	-0.54	0.058	61
-2.5	-0.02	1.000	83
-1.5	0.41	0.036	128
0.0	0.38**	0.006	142
1.5	0.23	0.205	128
2.5	-0.15	0.528	83
3.5	-0.26	0.400	61
4.5	0.19	0.712	41
5.5	0.24	0.778	30

2.6 Changing the Cut-point

One common design check involves imagining that the cut-point was at other locations and estimating what the treatment effect would be at those cut-points. Ideally, the estimated treatment effect would be clearest at the real cut-point. Table 7a shows the estimated treatment effect for every possible place that someone could put an imaginary cut-point. The real cut-point is 0, which is clearly the place where the effect is most evident. Thus, this test provides further evidence that the design worked as expected.

2.7 Addressing SUTVA Violations

The statistical tests carried out in the paper assume that each country in the sample has two potential outcomes following qualification, one if it went to the World Cup and one if it did not. Since there are 71 x 2 = 142 countries, the statistical tests assume that there are 142 x 2 = 284 total potential outcomes, exactly half of which we observe.

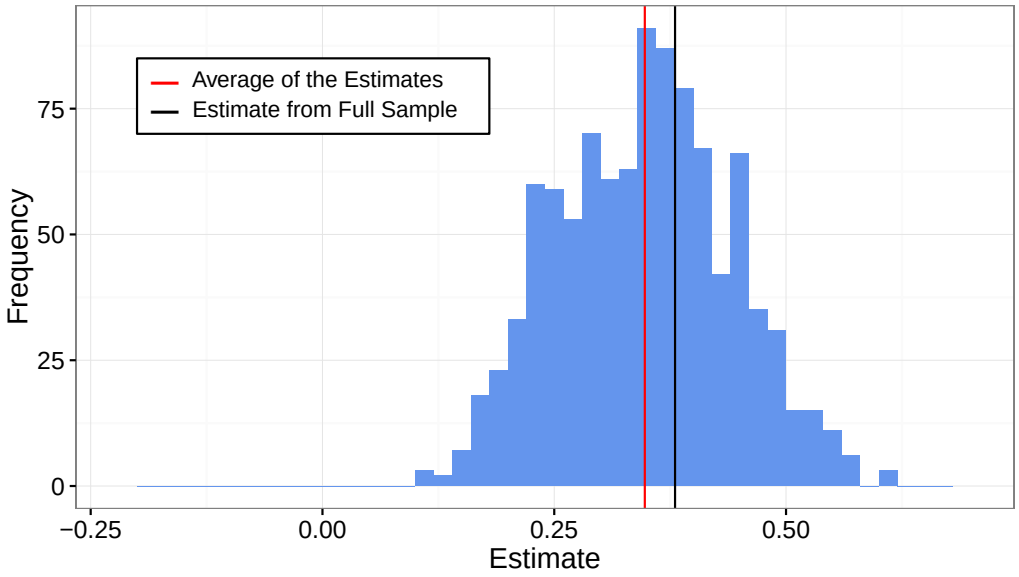
It would be a problem if the potential outcomes for a country changed when the treatment assignment of other countries changed. For instance, France could not have different potential outcomes depending on whether Japan was in the qualifier or non-qualifier group. If it did, then there would be more than 284 potential outcomes, and the statistical tests would not work. This requirement is called the Stable Unit Treatment Value Assumption (SUTVA). It is usually necessary for any experimental design to identify the treatment effect.

Fortunately, we do not need to worry about the treatment assignment of all countries across World Cup years, just the ones in our sample. These states are spread out geographically and over a long period of time. Thus, there is little reason to think that interference between countries raises a serious problem for this study.

However, there are some countries that appear in the sample multiple times. For instance, Yugoslavia is in the treatment group in 1958, and it shows up again in 1974. For these countries, the design requires that the treatment effect wears off by the time that the country reappears in the sample. Put simply, this condition is necessary for SUTVA to hold. Imagine that Yugoslavia's treatment assignment in 1958 influenced its potential outcomes in 1974. Then it would have 4 potential outcomes in 1974, 2 if it went to the World Cup in 1958 and 2 if it did not. Thus, there would be more than 284 potential outcomes across all the countries in the sample, and the assumptions behind the statistical tests would not be valid.

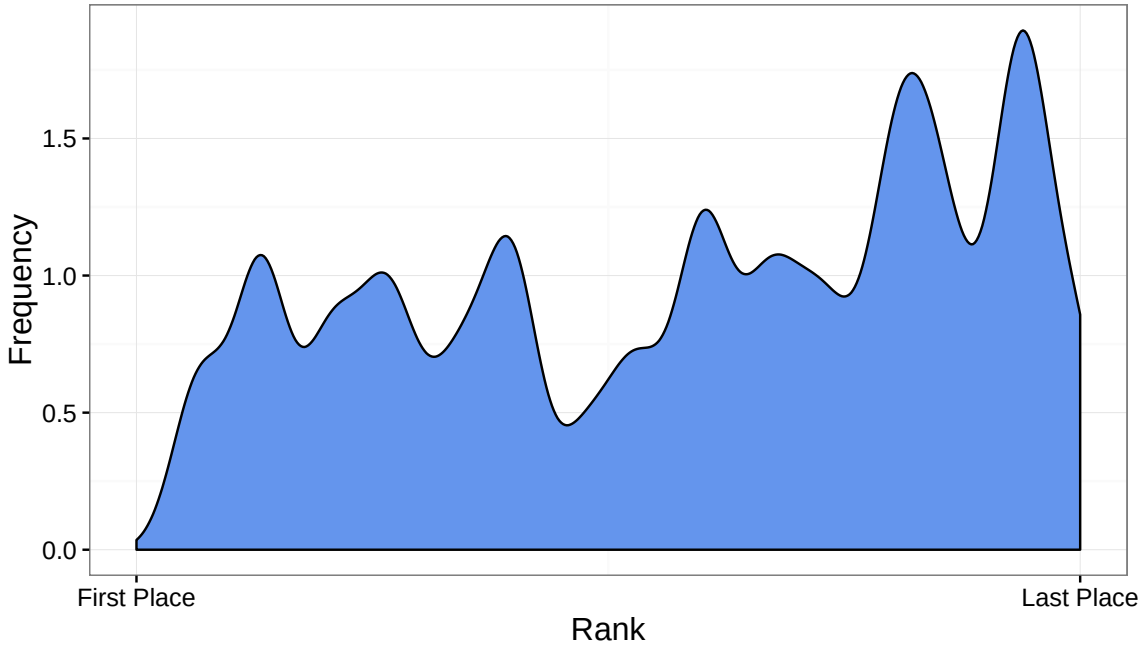
Although the time series graph (Figure 2) suggests that the treatment effect wears

Figure 4a. Estimates for Samples with No Repeated States



off before the fourth year following qualification, I carried out an additional robustness check to verify that repeat states were not influencing the results. Figure 4a shows the estimates for 1000 randomly chosen subsets of the World Cup sample where no country repeats. The sample size ranges from 31 to 39 pairs of countries. The estimated treatment effect is positive for all subsets, and the mean of these estimates is about equal to the average treatment effect. (Note: This procedure is similar to bootstrapping, with the exception that no country can be in the sample more than once.)

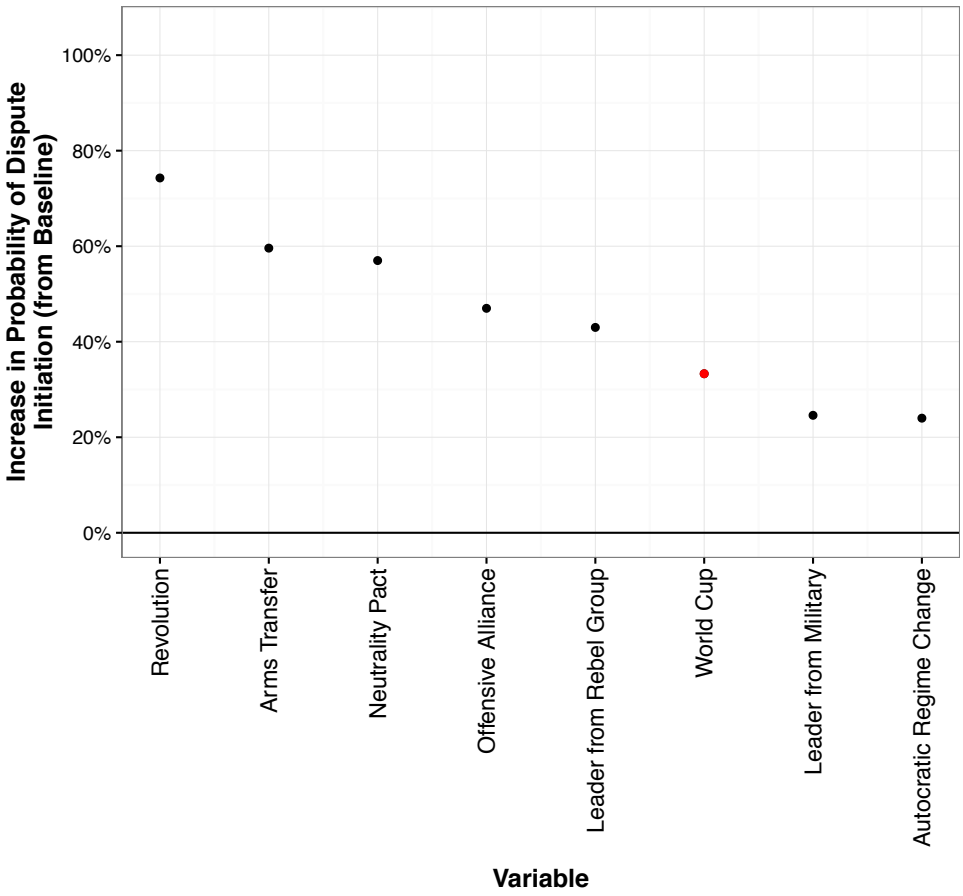
Figure 5a. Rankings of the Qualifiers at the World Cup



2.8 Performance of the Qualifiers at the World Cup

One possibility is that countries that barely qualify might tend to do poorly at the World Cup, so they would have a very different experience than most of the other countries that participated. This could mean that the effect found in our sample does not characterize how most countries respond to the World Cup. However, the countries in our sample actually did reasonably well at the World Cup. Figure 5a shows how they performed, as ranked by FIFA. The average scaled rank of the qualifiers in the sample was 0.59, which was only slightly worse than the average scaled rank for all participants (0.50). Thus, the experience of these countries at the World Cup was similar to the experience of most other participating states.

Figure 6a. Comparing the Effect of the World Cup to Other Estimated Treatment Effects



2.9 Putting the Effect Size in Perspective

Figure 6a shows how the estimated effect of the World Cup compares to other estimated treatment effects in the international relations literature. These estimated effects were taken from a number of recent articles that use the Militarized Interstate Dispute dataset (Enterline 1998:396; Leeds 2003:436; Krause 2004:365; Colgan 2010:682; Stam, Horowitz, and Ellis 2016:134). This comparison suggests that the impact of going to the World Cup is about two-fifths as large as a revolution, and that it is comparable to electing a leader with military experience. Thus, the World Cup is by no means the most powerful source of international conflict, but it is far from a trivial one.

Regional Championships Analysis

3.1 Constructing the Sample

I also collected data to test whether the regional soccer tournaments have a similar effect on state aggression. While these competitions are smaller, they still attract enormous audiences, especially within the participating countries. The states that compete in them are also closer to each other geographically. Thus, they are more likely to have existing disputes that could be exacerbated by nationalism from sports.

Fortunately, this hypothesis was not hard to test. These competitions are usually held every four years in North America, South America, Europe, Africa, Asia, and Oceania. Countries often must qualify for these tournaments in the same way that they do for the World Cup, so I was able to construct a new sample of countries that barely qualified and barely missed. This new sample consists of 78 countries, dating from 1960-2008. These countries are listed in Table 8a.

There is one difference between this sample and the World Cup sample. When constructing the World Cup sample, I used a two-point regression discontinuity window, which is optimal according to the method developed by Cattaneo, Titiunik, and Vazquez-Bare (2016). However, in the regional championship data, the optimal window size is one point. The reason is that the qualifiers and non-qualifiers are not as well balanced when the window is set at two points. I discuss this issue more in the next section.

Table 8a. Countries in the Regional Championship Sample

Qualifier	Non-qualifier	Year	Qualifier	Non-qualifier	Year
Israel	Iran	1960	Japan	Jordan	1988
Trinidad	Jamaica	1967	Zaire	Gabon	1992
Senegal	Guinea	1968	Britain	Ireland	1992
Taiwan	Japan	1968	Egypt	Tunisia	1992
El Salvador	Costa Rica	1977	Argentina	Brazil	1992
Belgium	Austria	1980	Liberia	Senegal	1996
Spain	Yugoslavia	1980	Thailand	Singapore	1996
Netherlands	Poland	1980	Kuwait	Lebanon	1996
Czechoslovakia	France	1980	Namibia	Gabon	1998
Greece	Hungary	1980	Algeria	Mali	1998
Haiti	Trinidad	1981	Vanuatu	Solomon Islands	1998
Portugal	Soviet Union	1984	Iran	Syria	2000
Denmark	Britain	1984	Qatar	Kazakhstan	2000
Romania	Sweden	1984	Tunisia	Gabon	2002
Spain	Netherlands	1984	Bahrain	Kuwait	2007
Syria	Indonesia	1984	Turkey	Norway	2008
India	Malaysia	1984	Russia	Britain	2008
Spain	Romania	1988	Netherlands	Bulgaria	2008
Denmark	Czechoslovakia	1988	South Africa	Uganda	2008
Ireland	Bulgaria	1988			

3.2 Checking for Balance

For the regional championship analysis, I use a one-point regression discontinuity window because this choice makes the balance between the qualifiers and non-qualifiers much better. Figure 7a illustrates this balance. As Figure 8a shows, balance for the two-point window is not nearly as good as we would want for reliable inference. However, while this balance is worse than would be expected in an experiment, it is important to note that the p-values do not indicate that there was sorting. Only two of the covariates are imbalanced at the 5% level, and only one other is imbalanced at the 10% level. Although most of the remaining p-values are below 0.5, this is partly because of correlations between these factors. The probability of seeing imbalance of this magnitude by chance across all the covariates is greater than 20%.

Moreover, if you use a two-point window for both the World Cup and regional championship samples and combine them (n=260), there is no evidence that countries sorted. The balance for this combined sample is shown in Figure 9a. If there was any systematic sorting, it should appear in this balance plot, since the sample size is large. However, the p-values appear to be distributed uniformly between 0 and 1. Thus, the fact that balance is less than ideal for countries that made or missed their regional championships by two points or less was probably caused by chance variation and is not indicative of a general sorting problem.

Figure 7a. Balance for the Regional Championships

Variable Name	Treatment Mean	Control Mean
Iron and Steel Production	8,669,210	8,862,260
Military Expenditures	4,508,910,000	9,970,380,000
Military Personnel	202,436	216,538
Total Population	44,140,100	29,555,800
Urban Population	11,988,500	10,780,500
Birth Rate	25.1	26.3
Death Rate	9.78	10.8
Infant Mortality	43.7	50.4
Energy Production	174,709,000	148,301,000
Imports	40,176,100,000	29,878,000,000
Exports	43,893,300,000	28,035,200,000
Land Area	428,727	446,295
Material Power Score	0.009	0.01
Level of Democracy	0.538	0.41
Great Power Status	0.05	0.103
Engaged in Civil War	0.05	0
Resolved Civil War	0	0
Year of State Formation	1886	1912
Sex Ratio	103	99.3
Life Expectancy	66.6	65.4
Median Age	26.1	25.6
Number of Alliances	16.3	14.9
U.S. Ally	0.462	0.256
Soccer Most Popular Sport	0.846	0.872
Appearance at Previous Regional Cup	0.436	0.385
MIDs Initiated in the Year Before	0.308	0.256
MIDs Initiated in the 3 Years Before	0.744	0.821
MIDs Initiated in the 5 Years Before	1.31	1.21

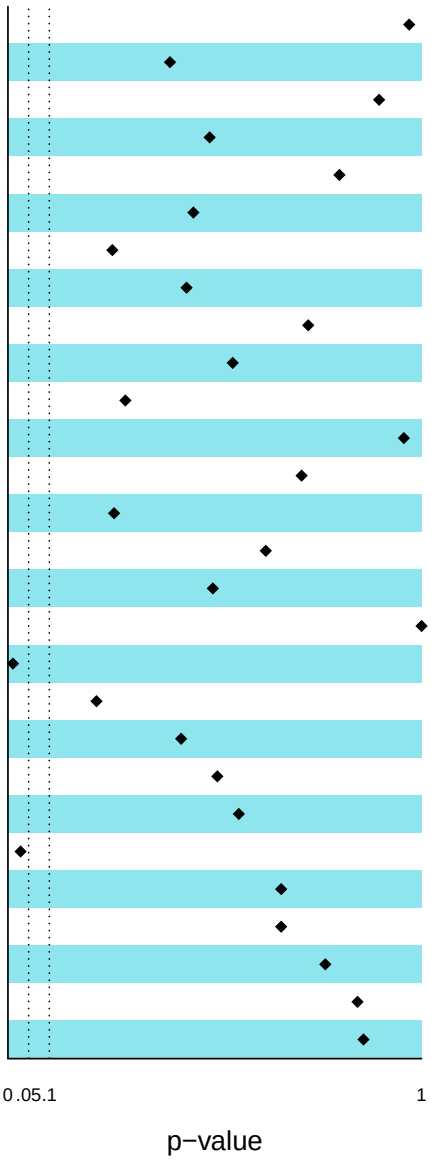


Figure 8a. Balance for the Regional Championships (Two-Point Window)

Variable Name	Treatment Mean	Control Mean
Iron and Steel Production	9,582,320	6,058,560
Military Expenditures	8,376,740,000	6,858,060,000
Military Personnel	291,305	183,153
Total Population	60,233,100	22,511,500
Urban Population	13,934,900	7,872,490
Birth Rate	28.1	30.2
Death Rate	10.7	11.7
Infant Mortality	55.5	60.8
Energy Production	197,757,000	107,942,000
Imports	33,015,600,000	22,767,300,000
Exports	35,726,500,000	20,814,200,000
Land Area	570,985	327,293
Material Power Score	0.01	0.008
Level of Democracy	0.441	0.305
Great Power Status	0.07	0.07
Engaged in Civil War	0.03	0
Resolved Civil War	0	0
Year of State Formation	1898	1924
Sex Ratio	101.6	99.9
Life Expectancy	63.7	62.5
Median Age	24.2	23.7
Number of Alliances	13.7	13.4
U.S. Ally	0.356	0.22
Soccer Most Popular Sport	0.831	0.864
Appearance at Previous World Cup	0.424	0.356
MIDs Initiated in the Year Before	0.305	0.22
MIDs Initiated in the 3 Years Before	0.797	0.661
MIDs Initiated in the 5 Years Before	1.41	0.949

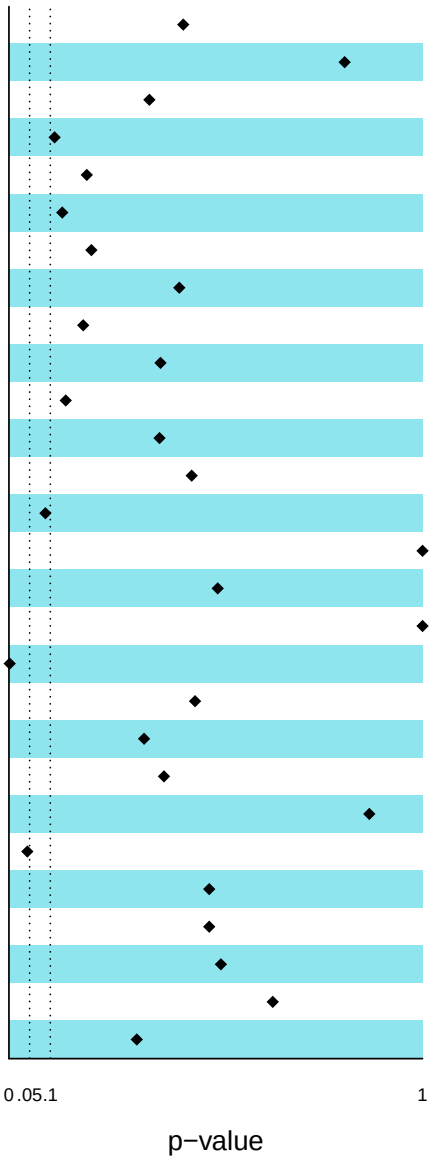
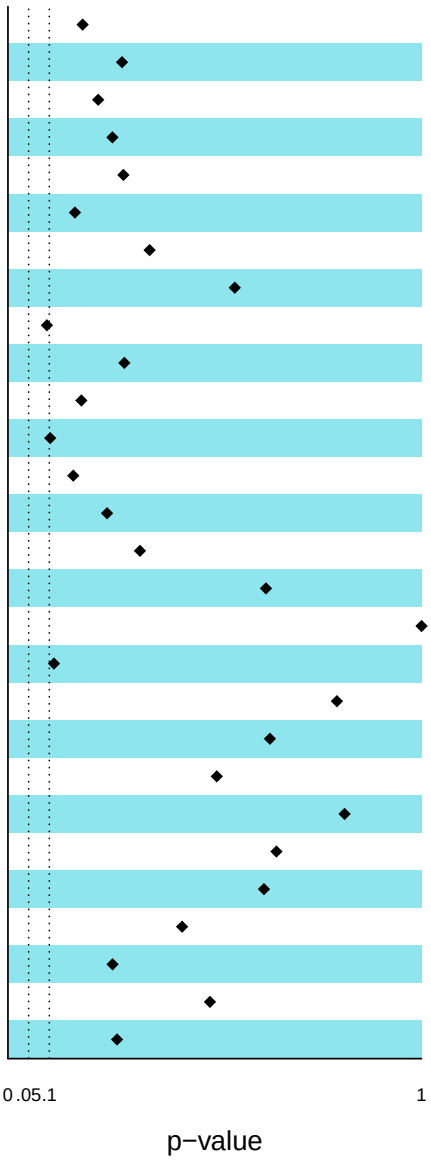


Figure 9a. Balance for the Combined Sample (Two-Point Window)

Variable Name	Treatment Mean	Control Mean
Iron and Steel Production	7,806,920	7,790,420
Military Expenditures	7,679,520,000	7,082,210,000
Military Personnel	272,131	233,038
Total Population	35,929,200	39,407,600
Urban Population	9,263,780	10,501,700
Birth Rate	24	28.5
Death Rate	11.1	11.2
Infant Mortality	51.5	56.9
Energy Production	151,756,000	145,444,000
Imports	22,059,500,000	26,328,400,000
Exports	21,282,900,000	26,546,500,000
Land Area	342,918	419,346
Material Power Score	0.01	0.009
Level of Democracy	0.446	0.4
Great Power Status	0.02	0.06
Engaged in Civil War	0.02	0.02
Resolved Civil War	0.008	0
Year of State Formation	1881	1906
Sex Ratio	97.9	100.3
Life Expectancy	65.6	63.6
Median Age	27.5	24.5
Number of Alliances	16.3	13.6
U.S. Ally	0.426	0.315
Soccer Most Popular Sport	0.931	0.854
Appearance at Previous World Cup	0.331	0.385
MIDs Initiated in the Year Before	0.154	0.246
MIDs Initiated in the 3 Years Before	0.646	0.708
MIDs Initiated in the 5 Years Before	0.992	1.14



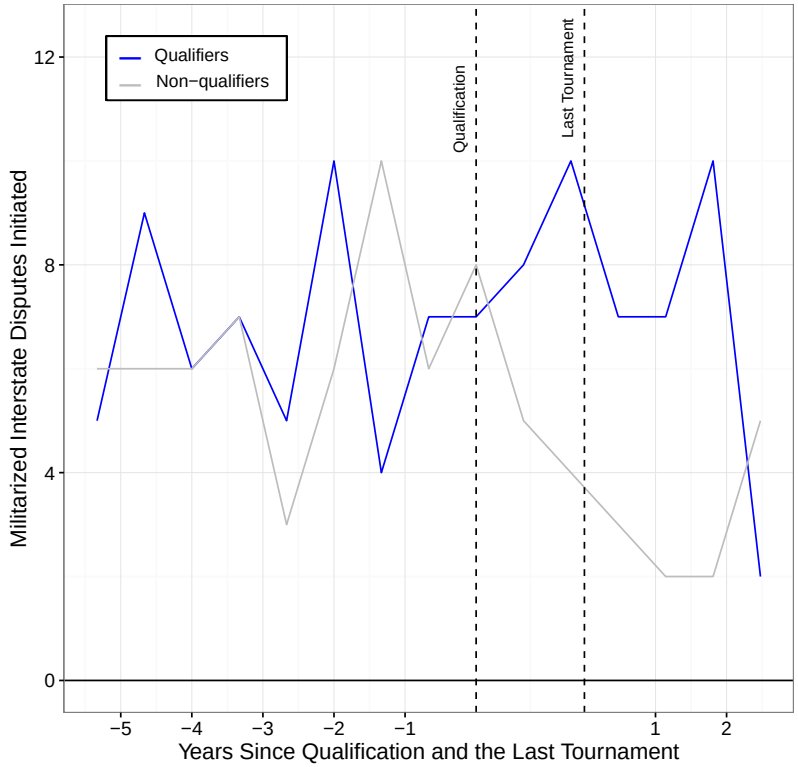
3.3 Additional Information About the Data

Between 1992 and 2010, the African tournament was held once every two years. To prevent overlap in the sample, I excluded several pairs of African countries because at least one member of the pair was in the previous tournament. These pairs include Senegal and Kenya (1993), Nigeria and Uganda (1993), Egypt and Morocco (1993), Zambia and Zimbabwe (1993), Guinea and Burundi (1993), Cameroon and Gabon (1993), Egypt and Senegal (1997), DRC and Liberia (1997), Mozambique and Malawi (1997), Congo and Mali (1999), Algeria and Liberia (1999), and Senegal and Zimbabwe (1999). Including these countries would have resulted in double counting the aggression levels for some countries in the years following qualification. Nevertheless, the difference between the aggression levels of the qualifiers and non-qualifiers following qualification remains significant when these pairs are included.

I also drop Iran and North Korea (1988) because Iran is an extreme outlier before qualification. It was coming out of the Iraq-Iran War and initiated 43 disputes in the three years before qualification. However, the difference between the aggression levels of the qualifiers and non-qualifiers following qualification remains significant when this dyad is included.

Also, because some of the tournaments did not happen for a year or more following qualification, I chose to look at the change in aggression in the three years before and the three years after qualification. These time intervals are 6 months longer than the time intervals used for the World Cup sample. However, I felt this was necessary given the amount of time between the qualification rounds and the regional tournaments.

Figure 10a. Aggression Before and After the Regional Championships

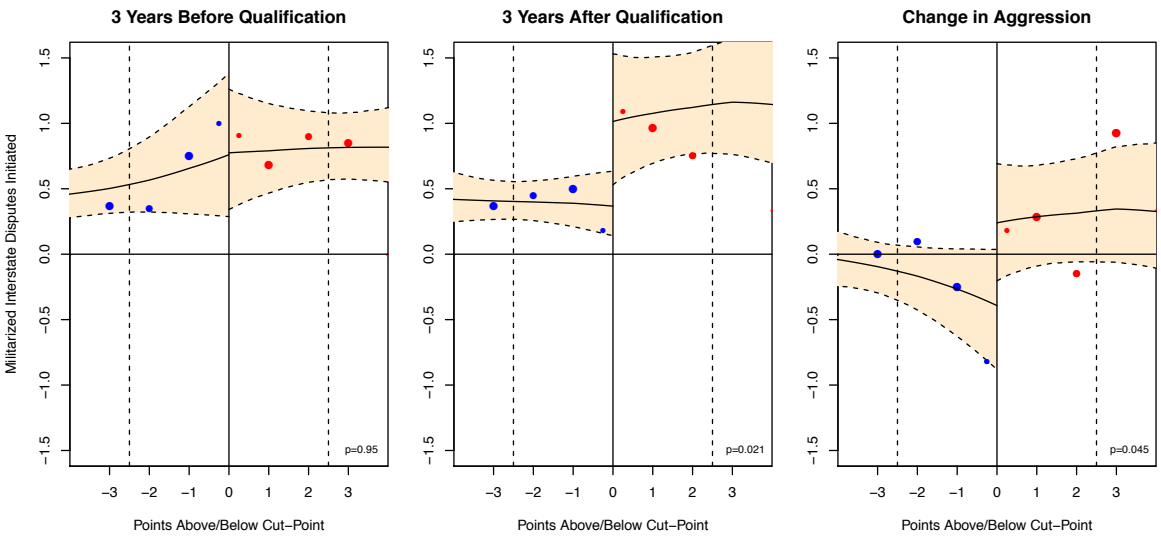


3.4 Tracking Aggression Levels

Figure 10a compares the aggression levels of the qualifiers and non-qualifiers before and after qualification. Much like with the World Cup sample, the separation between the two groups happens after qualification, and the treatment effect appears to wear off after the second year following the competitions. The results are significant for a two-tailed difference-in-differences test that compares the change in aggression between the qualifiers and the non-qualifiers before and after qualification (p=0.039).

There are several other similarities between these results and the ones presented in the paper. The disputes started by the qualifiers tended to be more revisionist and violent than the disputes started by the non-qualifiers. The median highest level of action for the qualifiers was 11.5, whereas the median for the non-qualifiers was 8. The results are also insensitive to a number of robustness checks, as I discuss in the following sections.

Figure 11a. Change in Aggression for the Regional Championships



3.5 Continuity Graph

In Figure 11a, I present the results using the continuity approach to regression discontinuity analysis. The graphs show that the qualifiers and non-qualifiers were well-balanced on aggression levels prior to qualification, but that the qualifiers experienced a substantial increase in aggression following qualification. The procedures used here are the same as the ones described in the paper.

3.6 Linear Regression

As before, I estimate the treatment effect using both difference-in-differences regression and controlling for the previous outcome. The results are not significant for the first approach, although the estimated treatment effect does not change much from the unadjusted estimate. However, the results are significant for all models using the second approach, and they do not come close to falling out of statistical significance. Note that the estimates in the two tables are similar and differ primarily in their standard errors. When the coefficient on the previous outcome is fixed at 1 (see Page 9), the standard errors will tend to be larger. Thus, the loss of significance for the first approach should not raise concern about the validity of the results.

Table 9a. Estimating the Effect of the Regional Championships on Dispute Initiation with Difference-in-Differences OLS Regression

	Model 1	Model 2	Model 3	Model 4
Regional Championship Appearance	0.67*	0.60	0.60	0.63
	(0.33)	(0.36)	(0.36)	(0.38)
Military Controls		X	X	X
Economic Controls			X	X
Demographic Controls				X

Table 10a. Estimating the Effect of the Regional Championships on Dispute Initiation Controlling for the Previous Outcome

	Model 1	Model 2	Model 3	Model 4
Regional Championship Appearance	0.62*	0.65*	0.67*	0.75*
	(0.25)	(0.26)	(0.26)	(0.28)
Military Controls		X	X	X
Economic Controls			X	X
Demographic Controls				X

Table 11a. Adjusting the Regression Discontinuity Window

Max Made/Missed	Estimated Effect	p-value	n
0 points	1.00	0.125	22
1 point	0.67*	0.039	78
2 points	0.36	0.167	118
3 points	0.53*	0.034	172
4 points	0.53*	0.030	178
5 points	0.46	0.057	188
6 points	0.52*	0.024	200
7 points	0.51*	0.021	206
8 points	0.47*	0.038	208
9 points	0.45*	0.042	212

3.7 Adjusting the Size of the Regression Discontinuity Window

As Table 11a shows, the results for the regional championship sample are largely insensitive to adjusting the regression discontinuity window. The estimated treatment effect is fairly stable, and the results are significant except when the window is set at zero, two, or five. The reason that the results are insignificant for the two-point window is that the countries that qualified by two points were much more aggressive than the countries that missed by two points before qualification (p=0.14). While they were much more aggressive after qualification as well, they experienced little change relative to the non-qualifiers. This issue is apparent from Figure 11a. However, it should not raise serious concern given that the results are significant for the one-point window, where balance between the qualifiers and non-qualifiers is much better.

Table 12a. Shifting the Minimum Score Requirement

Minimum Score	Estimated Effect	p-value	n
1 point	0.37	0.304	102
2 points	0.38	0.303	100
3 points	0.37	0.331	98
4 points	0.42	0.273	96
5 points	0.67*	0.039	78
6 points	0.65	0.069	68
7 points	0.56	0.122	64
8 points	0.54	0.232	52
9 points	0.57	0.244	42

3.8 Shifting the Minimum Score Requirement

Table 12a shows that the results for the regional championship sample are sensitive to shifting the minimum score requirement. Although the estimated treatment effect is pretty similar regardless of where the minimum score requirement is set, the results are only significant when it is fixed at five points. The reason that the results become insignificant when the window is set below five points is because the United States (1980) enters the non-qualifier group, and it experienced a large spike in aggression following qualification. If the dyad with the United States is removed, the results are significant if the minimum score requirement is set anywhere below five points. The fact that the results lose significance when the minimum score requirement is set above five points should not be concerning given the decreasing sample size.

Table 13a. Shifting the Location of the Cut-point

Location of Cut-point	Estimated Effect	p-value	n
-5.5	0.13	1.000	11
-4.5	0.20	0.610	8
-3.5	0.00	0.391	30
-2.5	0.10	0.727	47
-1.5	-0.35	0.433	48
0.0	0.67*	0.039	78
1.0	-0.44	0.346	48
2.5	1.08	0.197	47
3.5	-0.59	0.908	30
4.5	-1.33	0.464	8
5.5	2.17	0.108	11

3.9 Changing the Cut-point

Table 13a shows the estimated treatment effect for the regional championship sample at every possible place that someone could put an imaginary cut-point. As with the World Cup data, the real cut-point at 0 is clearly the place where the effect is most evident. So this test provides again confirms that the design worked as expected.

Combined Sample

4.1 Balance Plot

Figure 12a shows the balance between the qualifiers and non-qualifiers in the combined World Cup and regional championship samples (n=220). None of the differences are significant at even the 10% level, so this balance is much better than would be expected in an experiment. Given the large sample size for this test, there is little reason to suspect that the countries considered in this study sorted at the cut-point. Moreover, the results are significant at the 0.1% level for the combined sample (p=0.0006). In other words, the probability of seeing these trends in both datasets by chance is about 1/1400.

Figure 12a. Balance for the Combined Sample

Variable Name	Treatment Mean	Control Mean
Iron and Steel Production	7,293,350	8,966,350
Military Expenditures	8,042,010,000	6,365,850,000
Military Personnel	246,518	244,564
Total Population	27,009,200	43,772,600
Urban Population	7,637,250	12,097,900
Birth Rate	23.8	25.2
Death Rate	11	10.6
Infant Mortality	49.4	49.6
Energy Production	144,425,000	157,368,000
Imports	21,802,000,000	30,301,100,000
Exports	19,951,500,000	31,749,700,000
Land Area	339,209	390,080
Material Power Score	0.01	0.01
Level of Democracy	0.491	0.445
Great Power Status	0.03	0.05
Engaged in Civil War	0.009	0.03
Resolved Civil War	0.009	0
Year of State Formation	1881	1891
Sex Ratio	97.8	100.1
Life Expectancy	66	65.8
Median Age	27.7	26.4
Number of Alliances	16.4	15.4
U.S. Ally	0.44	0.382
Soccer Most Popular Sport	0.927	0.882
Appearance at Previous World Cup	0.327	0.391
MIDs Initiated in the Year Before	0.173	0.218
MIDs Initiated in the 3 Years Before	0.682	0.691
MIDs Initiated in the 5 Years Before	1.04	1.11

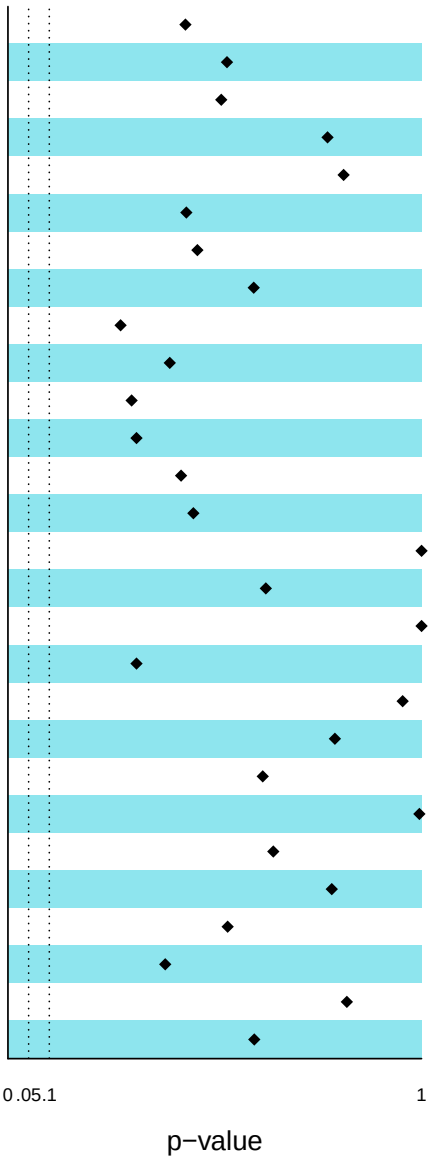
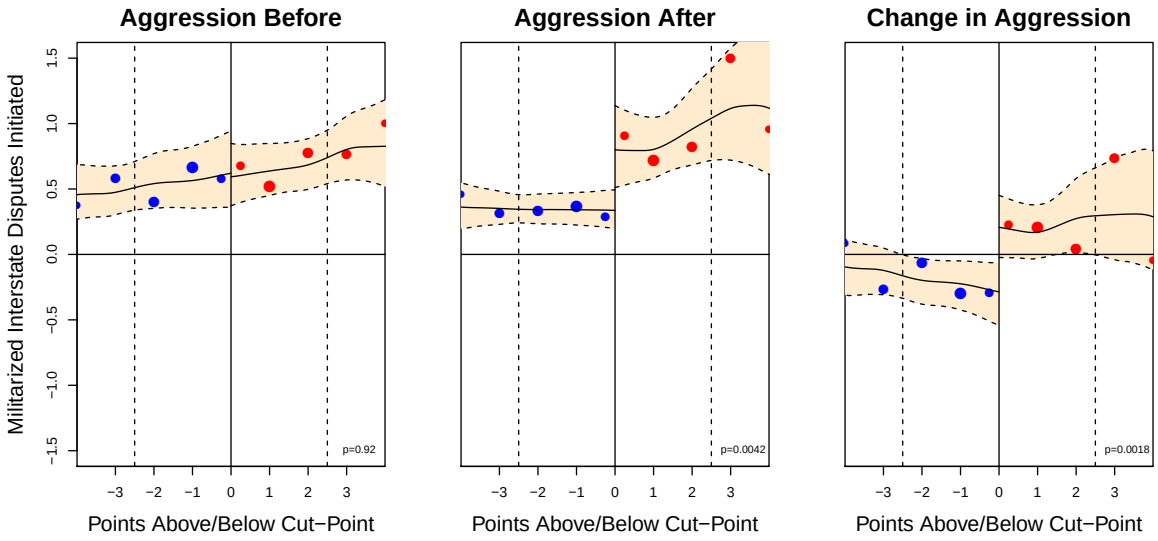


Figure 13a. Change in Aggression for the Combined Sample



4.2 Continuity Graph

Figure 13a shows the results for the continuity approach. As the graph on the left shows, there is no discontinuity at the cut-point in the years leading up to qualification. On the other hand, there is a large discontinuity at the cut-point in the years following qualification, as shown in the middle and right graphs. The results for both of these graphs are significant at the 1% level.

Information About the Covariates

The following variables were taken from the Correlates of War Database (Ghosn, Palmer, Bremer 2004). Iron and Steel Production (in tons), Military Expenditures (in current year U.S. dollars), Military Personnel, Energy Production (in coal-ton equivalents), Total Population, and Urban Population. The Material Power Score is a weighted average of these variables, also known as the COW CINC Score. The database is available online at http://www.correlatesofwar.org/.

The remaining data was pieced together from a number of sources. The demographic data is from the United Nations World Population Prospects Survey. This includes Birth Rate, Death Rate, Sex Ratio (Male:Female), Infant Mortality, Life Expectancy, and Median Age. The database is available at http://esa.un.org/wpp/. The democracy data was taken from the Polity IV database (Marshall, Gurr, and Jaggers 2013), archived here: http://www.systemicpeace.org/polity/polity4.htm.

I do not include gross domestic product in the balance plot because there is not a dataset that provides reliable measurements of this variable for all of the countries considered in this study. In fact, even the most comprehensive datasets like the Penn World Tables and World Development Indicators fall well short of providing all of the necessary information. However, many of the economic and demographic factors included in this study are related to GDP, especially some of the production and population figures, as well as Birth Rate and Infant Mortality.

References

Cattaneo, Matias D., Rocio Titiunik, and Gonzalo Vazquez-Bare. 2016. "Inference in Regression Discontinuity Designs under Local Randomization." *Stata Journal* 16(2): 331-367.

Colgan, Jeff D. 2010. "Oil and Revolutionary Governments: Fuel for International Conflict." *International Organization* 64(4): 661-694.

Enterline, Andrew J. 1998. "Regime Changes and Interstate Conflict, 1816-1992." *Political Research Quarterly* 51(2): 385-409.

Ghosn, F., G. Palmer, and S. Bremer. 2004. "The MID3 Data Set, 1993-2001: Procedures, Coding Rules, and Description." *Conflict Management and Peace Science* 21: 133-154.

Gowa, Joanne. 1998. "Politics at the Water's Edge: Parties, Voters, and the Use of Force Abroad." *International Organization* 52(2): 307-324.

Krause, Volker. 2004. "Hazardous Weapons? Effects of Arms Transfers and Defense Pacts on Militarized Disputes, 1950-1995." *International Interactions* 30(4): 349-371.

Lai, Brian, and Dan Slater. 2006. "Institutions of the Offensive: Domestic Sources of Dispute Initiation in Authoritarian Regimes, 1950-1992." American Journal of Political Science 50(1): 113-126.

Leeds, Brett Ashley. 2003. "Do Alliances Deter Aggression? The Influence of Military Alliances on the Initiation of Militarized Interstate Disputes." *American Journal of Political Science* 47(3): 427-439.

Marshall, M., T. Gurr, and K. Jaggers. 2013. *POLITY IV PROJECT: Political Regime Characteristics and Transitions, 1800-2012. Dataset Users Manual*. Center for Systemic Peace. Dataset version p4v2012. www.systemicpeace.org.

Stam, Allan C., Michael C. Horowitz, and Cali M. Ellis. 2015. *Why Leaders Fight*. New York: Cambridge University Press.

United Nations, Department of Economic and Social Affairs, Population Division. 2013. *World Population Prospects: The 2012 Revision, Key Findings and Advance Tables*. Working Paper No. ESA/P/WP.227.