INTRODUCING THE WARRING-STATES JAPAN BATTLE DATA:

APPENDIX

This appendix presents more detailed descriptive statistics for the "Warring-States Japan Battle Data," as well as the details of the robustness tests carried out in the article.

Descriptive Statistics

Table A1, below, lists a few of the more notable and well-known battles and wars of the Warring-States period, and their dates and unique IDs in the battle data.

Table A1: Notable Battles & Wars, 1467-1600

Date	Battle	battle_id
1467/1/18	Outbreak of the Ōnin War	1
1493/[4]/22	The Coup of Meiou	441
1495/9	Hōjō Conquest of Odawara Castle	485
1546/4/20	Night Battle of Kawagoe	1297
1551/8/27	Tainei-ji Incident	1403
1553/8	First Battle of Kawanaka-jima	1451
1555/7/19	Second Battle of Kawanaka-jima	1497
1555/10/1	Battle of Itsuku-shima	1506
1557/8	Third Battle of Kawanaka-jima	1554
1560/5/19	Battle of Okehazama	1607
1561/9/10	Fourth Battle of Kawanaka-jima	1649
1564/8/3	Fifth Battle of Kawanaka-jima	1754
1571/9/12	The Siege of Mount Hiei	2034
1572/12/22	Battle of Mikatagahara	2070
1575/5/21	Battle of Nagashino	2186
1577/10/10	Battle of Shigisan Castle	2268
1580/4/9	Siege of Ishiyama Hongan-ji	2386
1582/6/2	The Honnoh-ji Incident	2489
1583/4/21	Battle of Shizugatake	2558
1584/4/9	Battle of Owari-Nagakute	2603
1587/4/17	Battle of Nejiro-zaka	2728
1590/4/1	Seige of Odawara	2785
1600/9/15	Battle of Sekigahara	2877

Table A2, below, is a list of the number of individual battles observed in each of Japan's 68 premodern provinces from 1467-1600.

Table A2: Battles by Province, 1467-1600

Yamashiro (山城)	195	Kōzuke (上野)	46	Dewa (出羽)	17
Settsu (摂津)	123	Ōsumi (大隅)	45	Hōki (伯耆)	15
Shinano (信濃)	120	Izumi (和泉)	44	Sanuki (讃岐)	15
Mutsu (陸奥)	117	Owari (尾張)	42	Tango (丹後)	15
Hizen (肥前)	114	Shimotsuke (下野)	40	Noto (能登)	14
Ōmi (近江)	113	Tanba (丹波)	38	Awa2 (阿波)	13
Yamato (大和)	104	Hitachi (常陸)	37	Izu (伊豆)	13
Chikuzen (筑前)	85	Kai (甲斐)	37	Mimasaka (美作)	13
Aki (安芸)	82	Satsuma (薩摩)	35	Inaba (因幡)	12
Echigo (越後)	70	Tosa (土佐)	35	Awa1 (安房)	7
Hyūga (日向)	69	Bingo (備後)	33	Hida (飛騨)	7
Kawachi (河内)	69	Iwami (石見)	30	Kazusa (上総)	7
Musashi (武蔵)	64	Kaga (加賀)	30	Iga (伊賀)	6
Mikawa (三河)	61	Sagami (相模)	30	Tajima (但馬)	6
Higo (肥後)	58	Kii (紀伊)	29	Iki (壱岐)	5
Mino (美濃)	58	Shimōsa (下総)	26	Awaji (淡路)	4
Harima (播磨)	55	Suruga (駿河)	26	Nagato (長門)	4
Buzen (豊前)	53	Chikugo (筑後)	25	Tsushima (対馬)	4
Tōtōmi (遠江)	53	Bizen (備前)	21	Oki (隠岐)	3
Echizen (越前)	51	Bungo (豊後)	21	Wakasa (若狭)	3
Ise (伊勢)	50	Bicchū (備中)	19	Sado (佐渡)	2
Ecchū (越中)	49	Suō (周防)	19	Shima (志摩)	0
Izumo (出雲)	47	Iyo (伊予)	18	Unknown	124

Figure A1, below, presents the battle by province data in cartographic form.

Figure A1: Battles by Province (Map), 1467-1600

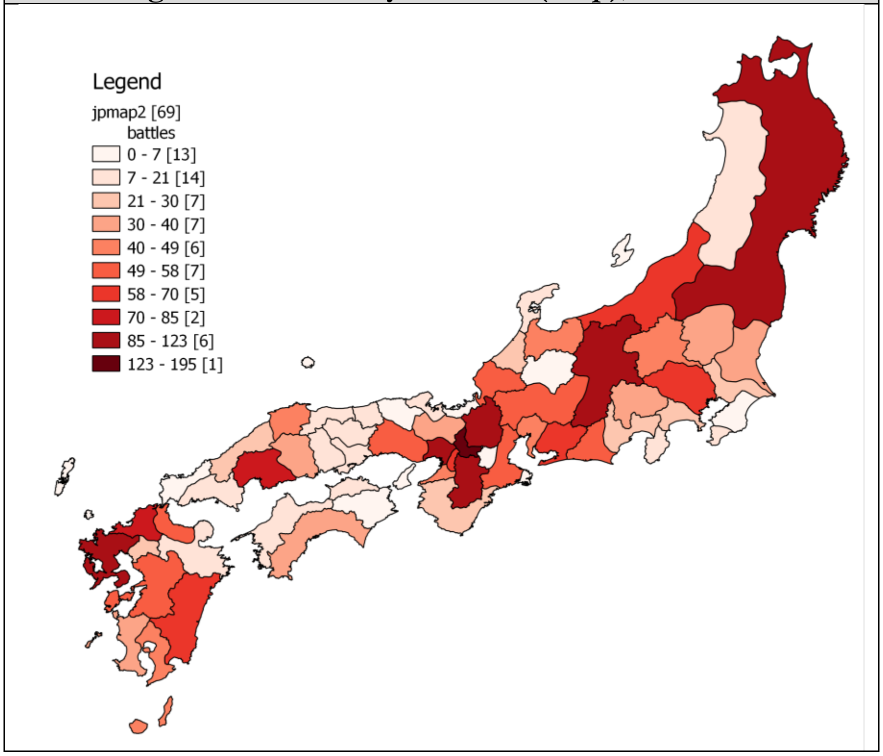
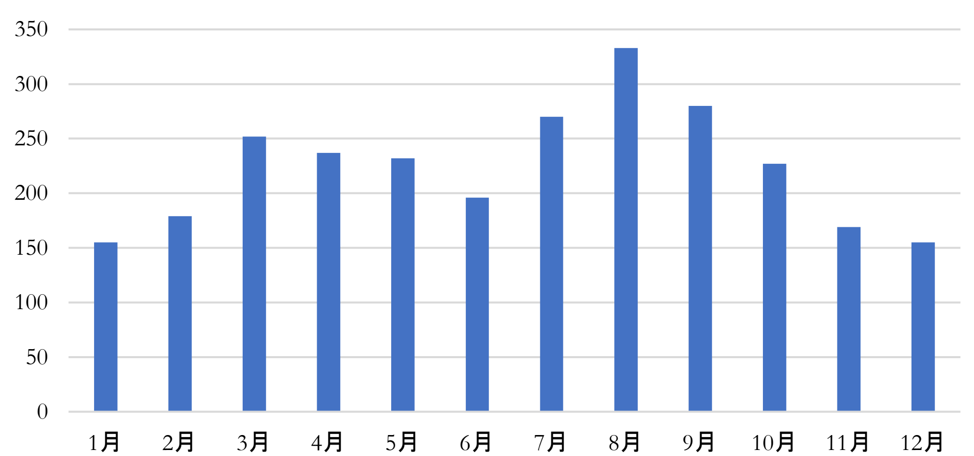


Figure A2, below, presents the number of battles for each of the twelve months of the year.

Figure A2: Battles by Month, 1467-1600



Note, however, that Figure A2 presents battles by month according to Japan's traditional lunisolar calendar. This means that periodic "leap months" are not included in the figure.[1] The Warring-States Japan Battle Data also include globally-standard "Gregorian" calendar dates, which are presented, along with the lunisolar dates, in Table A3 (below).

Table A3: Battles by Month for Lunisolar and Gregorian Calendars, 1467-1600

Month	Battles (Lunisolar Calendar)	Battles (Gregorian Calendar)
1 月	155	127
2 月	179	140
3 月	252	153
4 月	237	197
5 月	232	242
6 月	196	198
7 月	270	178
8 月	333	269
9 月	280	290
10 月	227	256
11 月	169	219
12 月	155	184

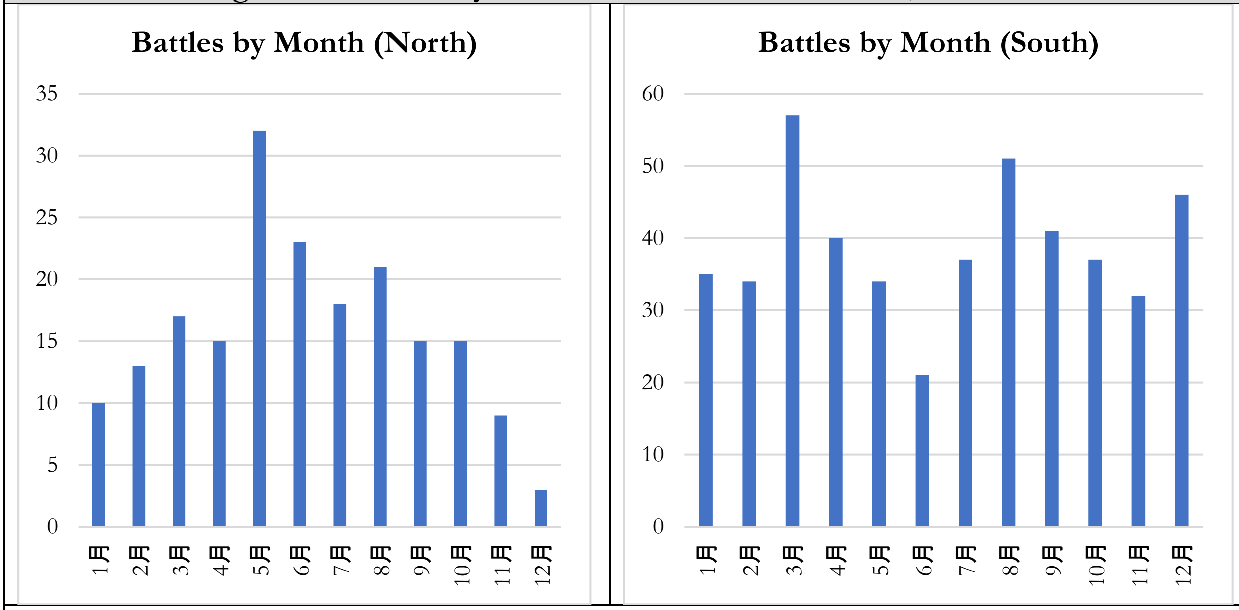
Note: "Leap months" are excluded for Lunisolar Calendar data; seasons and ranges of months are excluded for both. There are fewer aggregate battles in the Gregorian column because the conversion process created more battles with ranges of months, which are not included in the table. See the Codebook for details.

As noted in the article, there is also interesting seasonal variation depending on the region of the country we examine. Figure A3, below, shows the distribution of battles by month in Japan's four northernmost provinces (Dewa, Mutsu, Echigo, Sado) on the left and in Japan's southernmost provinces (the eleven provinces on Kyushu[2]) on the right. As the figures make clear, there is much greater variation in the seasonal distribution of battles in the north than in the south, due to the harsh winters experienced in northern Japan. And the apparent influence of the monsoon rains on battle in the south (see the sixth month) is considerably greater than in the north, which is in accord with contemporary weather patterns in Japan.

[1] The figure also excludes battles only specified to seasons (e.g., those taking place in "summer" or "fall") and battles only specified to ranges of months (e.g., those taking place in months 6~7).

[2] These are: Bungo, Buzen, Chikugo, Chikuzen, Higo, Hizen, Hyūga, Iki, Ōsumi, Satsuma, Tsushima.

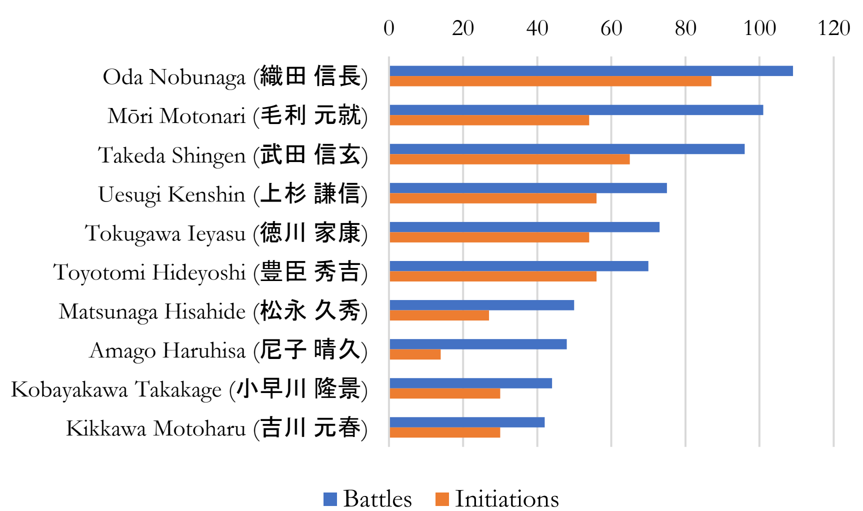
Figure A3: Battles by Month in the North and South, 1467-1600



Note: this figure presents lunisolar calendar dates, and therefore leap months, ranges of months, and seasons are excluded from the data.

Figure A4, below, presents the ten individual army leaders with the greatest number of battles. The figure also presents the number of initiations for each of these belligerents, and the combination of these two values provides something of an "aggressiveness ratio" for each belligerent, with some (Oda, Uesugi, Toyotomi) attacking at a much higher rate than others (Mōri, Amago).

Figure A4: Battles by Belligerent, 1467-1600



The Diffusion of Conflict

This section presents some additional details on the diffusion analysis in the article, as well as full results for the robustness tests.

As noted in the article, I use the province as the spatial unit and the year as the temporal unit, making the province-year the unit of analysis. The structure of the data is therefore monadic. Some scholars have used dyadic data to examine the diffusion of conflict (Black 2013; Forsberg 2014, 195-6), so, as a robustness test, I analyze a dyadic version of the data (see Table A12, below).

As also noted in the article, I use dichotomous variables for both the independent (IV) and dependent variables (DV). The two key IVs indicate whether at least one battle took place in a territorially-proximate province in the recent past (*adjacent_battle*) and whether at least one battle took place in a territorially-proximate province that *shares a major road* with the province in question in the recent past (*road_battle*). Therefore, for both of these IVs, even if multiple proximate provinces experience battle in the recent past, or there are multiple battles within a proximate province, it still only counts as a single, dichotomous observation. One concern might be that this is an overly-coarse measure that "throws away" a lot of useful data. To address such concerns, I rerun the analysis using dyadic data with numerical count variables for battle rather than dichotomous variables as a robustness test (see Table A12, below).

As noted in the article, I define these IVs somewhat conservatively, considering only provinces that share a land border with the province in question as "territorially-proximate," and only battles that took place in the same calendar year (though, of course, at an earlier date) or the prior calendar year as "recent." What this means is that the shortest possible duration between "recent" battles is a single day (e.g., 1/1/1467 & 1/2/1467) and the longest possible duration is just shy of two years (e.g., 1/1/1467 & 12/30/1468).

As noted in the article, premodern Japan had a fairly extensive road network, which was completed over the course of the seventh to ninth centuries. The existence of this network is the key to what differentiates the two primary IVs: *adjacent_battle* and *road_battle*. *Adjacent_battle* simply indicates whether a province sharing a land border with a given province experienced conflict within the past 24 months. *Road_battle* adds the condition that this adjacent province also shares a major road with the province in question. The map that was used as the source for the road network data (i.e., indicating which provinces were connected by road to which other provinces) is available online through the JapanKnowledge encyclopedia.[3]

The analysis also includes a number of control variables. Table A4, below, presents summary statistics for the DV (*battle*), the two key IVs (*adjacent_battle*, *road_battle*), and the control variables.

Table A4: Summary Statistics

Variable	Mean	SD	Min	Max
battle	0.1598	0.3664	0	1
adjacent_battle	0.5961	0.4907	0	1
road_battle	0.4485	0.4974	0	1
neighbors	4.032	1.642	1	10
area	4,531.6	6338.4	343.4	48,432.1
terrain_ruggedness	203.7	81.36	15.3	394.2

Note: there is no missing data on any variable.

[3] See: "五畿七道" in JapanKnowledge 2022.

Analysis

The main analysis was conducted using a linear probability model with robust standard errors and two-way fixed effects for province and year. Table A5, below, presents the main results with a little more detail than in the article.

Table A5: Linear Probability Analysis

	Dependent variable: battle		
	(1)	(2)	(3)
adjacent_battle	0.018*	-0.007	
	(0.009)	(0.013)	
road_battle		0.033***	0.029***
		(0.013)	(0.009)
neighbors	0.137***	0.135***	0.135***
	(0.020)	(0.019)	(0.019)
area	0.0001***	0.0001***	0.0001***
	(0.00001)	(0.00001)	(0.00001)
terrain_ruggedness	-0.001***	-0.001***	-0.001***
	(0.0004)	(0.0004)	(0.0004)
Constant	-0.153	-0.146	-0.146
	(0.160)	(0.160)	(0.160)
Observations	8,442	8,442	8,442
R2	0.136	0.136	0.136
Adjusted R2	0.115	0.116	0.116
Residual Std. Error	0.345 (df = 8245)	0.345 (df = 8244)	0.345 (df = 8245)
F Statistic	6.603*** (df = 196; 8245)	6.611*** (df = 197; 8244)	6.644*** (df = 196; 8245)
Note:			*p<0.1; **p<0.05; ***p<0.01

Robustness Tests

I conducted a number of robustness tests to see how sensitive the results are to alternative model specifications and the inclusion of alternative and additional controls. First, I reran the analysis using Logistic regression rather than linear regression. Here, again, I include robust standard errors and two-way fixed effects for province and year. Table A6, below, presents the results. The results are similar, though, here the coefficient on *road_battle* is only statistically significant at the 0.1 level in model 2 and at the 0.05 level in model 3.

Table A6: Logistic Regression Analysis

	Dependent variable: battle		
	(1)	(2)	(3)
adjacent_battle	0.152*	0.001	
	(0.088)	(0.123)	
road_battle		0.192*	0.193**
		(0.108)	(0.078)
neighbors	1.087***	1.075***	1.075***
	(0.178)	(0.178)	(0.178)
area	0.001***	0.001***	0.001***
	(0.0002)	(0.0002)	(0.0002)
terrain_ruggedness	-0.013***	-0.013***	-0.013***
	(0.003)	(0.003)	(0.003)
Constant	-5.054***	-5.011***	-5.011***
	(1.128)	(1.126)	(1.126)
Observations	8,442	8,442	8,442
Log Likelihood	-3,091.359	-3,089.666	-3,089.666
Akaike Inf. Crit.	6,576.719	6,575.332	6,573.332
Note:			*p<0.1; **p<0.05; ***p<0.01

Second, I reran the analysis using a linear probability model, the same controls, and with robust standard errors, but with a lagged dependent variable (*lagged_battle*) rather than two-way fixed effects. The results are presented in Table A7, below. Here the results are unchanged, though the coefficients on *road_battle* in models 2 and 3 are substantively larger (0.033 → 0.052 and 0.29 → 0.58).

Table A7: Linear Probability Analysis with Lagged DV

	Dependent variable: battle		
	(1)	(2)	(3)
adjacent_battle	0.046***	0.008	
	(0.008)	(0.011)	
road_battle		0.052***	0.058***
		(0.012)	(0.008)
lagged_battle	0.308***	0.304***	0.305***
	(0.014)	(0.014)	(0.014)
neighbors	0.013***	0.013***	0.013***
	(0.003)	(0.003)	(0.003)
area	0.00000***	0.00000***	0.00000***
	(0.00000)	(0.00000)	(0.00000)
terrain_ruggedness	-0.0002***	-0.0001***	-0.0001***
	(0.00005)	(0.00005)	(0.00005)
Constant	0.051***	0.046***	0.047***
	(0.011)	(0.011)	(0.012)
Observations	8,379	8,379	8,379
R2	0.123	0.125	0.125
Adjusted R2	0.122	0.124	0.124
Residual Std. Error	0.343 (df = 8373)	0.343 (df = 8372)	0.343 (df = 8373)
F Statistic	233.993*** (df = 5; 8373)	198.973*** (df = 6; 8372)	238.701*** (df = 5; 8373)
Note:			*p<0.1; **p<0.05; ***p<0.01

Third, I combined the previous two robustness tests and ran a Logistic regression model, with the same controls, and with robust standard errors, but with a lagged dependent variable rather than two-way fixed effects. Table A8, below, presents the results. The results are unchanged, though here the coefficient on *road_battle* in model 2 has increased in statistical significance to the 0.01 level as compared to the Logistic model with two-way fixed effects (Table A6, above).

Table A8: Logistic Regression Analysis with Lagged DV

	Dependent variable: battle		
	(1)	(2)	(3)
adjacent_battle	0.459***	0.140	
	(0.074)	(0.109)	
road_battle		0.408***	0.501***
		(0.099)	(0.067)
lagged_battle	1.717***	1.692***	1.699***
	(0.070)	(0.071)	(0.070)
neighbors	0.105***	0.103***	0.108***
	(0.021)	(0.021)	(0.020)
area	0.00002***	0.00002***	0.00002***
	(0.00000)	(0.00000)	(0.00000)
terrain_ruggedness	-0.001***	-0.001***	-0.001***
	(0.0004)	(0.0004)	(0.0004)
Constant	-2.638***	-2.680***	-2.662***
	(0.106)	(0.107)	(0.106)
Observations	8,379	8,379	8,379
Log Likelihood	-3,245.526	-3,236.384	-3,237.216
Akaike Inf. Crit.	6,503.052	6,486.769	6,486.432
Note:	*p<0.1; **p<0.05; ***p<0.01		

Fourth, I reran the analysis using a linear probability model, the same controls, and with robust standard errors, but with *peace_years* rather than two-way fixed effects. *Peace_years* is a simple count variable of the number of years since a given province last observed battle.[4] As shown in Table A9 (below), the main results are similar, though the coefficient on *adjacent_battle* is now statistically significant at the 0.01 level in models 1 and 2 (though substantively smaller than that of *road_battle* in model 2) and the coefficients on *road_battle* are substantively larger in models 2 and 3 than in the main analysis (0.033 → 0.054 and 0.29 → 0.74).

Table A9: Linear Probability Analysis with Peace_years

	Dependent variable: battle		
	(1)	(2)	(3)
adjacent_battle	0.068***	0.028***	
	(0.008)	(0.011)	
road_battle		0.054***	0.074***
		(0.011)	(0.008)
peace_years	-0.006***	-0.006***	-0.006***
	(0.0002)	(0.0002)	(0.0002)
neighbors	0.004	0.004	0.005*
	(0.003)	(0.003)	(0.003)
area	0.00000***	0.00000***	0.00000***
	(0.00000)	(0.00000)	(0.00000)
terrain_ruggedness	-0.0001**	-0.0001*	-0.0001*
	(0.00005)	(0.00005)	(0.00005)
Constant	0.193***	0.185***	0.186***
	(0.013)	(0.013)	(0.013)
Observations	8,442	8,442	8,442
R2	0.109	0.112	0.111
Adjusted R2	0.109	0.111	0.110
Residual Std. Error	0.346 (df = 8436)	0.346 (df = 8435)	0.346 (df = 8436)
F Statistic	206.910*** (df = 5; 8436)	176.526*** (df = 6; 8435)	210.568*** (df = 5; 8436)
Note:			*p<0.1; **p<0.05; ***p<0.01

[4] In constructing this variable, I set all provinces to 1 in 1467 and began counting upward each year until a province's first battle, at which point *peace_years* was set to 0, and it began the count again. If a province experienced battle in 1467, it started at 0.

Fifth, I reran the linear probability analysis with the same controls, with robust standard errors, and with two-way fixed effects, but adding a dichotomous variable controlling for each of Japan's eight premodern regions, one-by-one, to see if any was importantly driving the results. Regions are defined by the ancient Goki Shichidō (five home provinces, seven circuits) system.[5] For each of these analyses, only the "model 2" specifications above (i.e., including *adjacent_battle* and *road_battle*) are included. Tables A10 and A11 present the results. As is clear in these tables, the results are unchanged.

Table A10: Linear Probability Analysis with Region Controls (1)

	Dependent variable: battle			
	(1)	(2)	(3)	(4)
adjacent_battle	-0.007	-0.007	-0.007	-0.007
	(0.013)	(0.013)	(0.013)	(0.013)
road_battle	0.033***	0.033***	0.033***	0.033***
	(0.013)	(0.013)	(0.013)	(0.013)
neighbors	0.135***	0.135***	0.104***	-0.005
	(0.019)	(0.019)	(0.022)	(0.099)
area	0.0001***	0.0001***	0.00004	0.00004
	(0.00001)	(0.00001)	(0.00003)	(0.00004)
terrain_ruggedness	-0.001***	-0.001***	-0.001	-0.002***
	(0.0004)	(0.0004)	(0.001)	(0.0003)
region_tousan	-1.357***			
	(0.193)			
region_toukai		0.057		
		(0.044)		
region_hokuriku			-0.129	
			(0.099)	
region_kinai				0.266
				(0.205)
Constant	-0.146	-0.146	0.010	0.499
	(0.160)	(0.160)	(0.144)	(0.439)
Observations	8,442	8,442	8,442	8,442
R2	0.136	0.136	0.136	0.136
Adjusted R2	0.116	0.116	0.116	0.116
Residual Std. Error (df = 8244)	0.345	0.345	0.345	0.345
F Statistic (df = 197; 8244)	6.611***	6.611***	6.611***	6.611***
Note:	*p<0.1; **p<0.05; ***p<0.01			

[5] See: "Goki Shichidō," in JapanKnowledge 2022.

Table A11: Linear Probability Analysis with Region Controls (2)

	Dependent variable: battle			
	(1)	(2)	(3)	(4)
adjacent_battle	-0.007	-0.007	-0.007	-0.007
	(0.013)	(0.013)	(0.013)	(0.013)
road_battle	0.033***	0.033***	0.033***	0.033***
	(0.013)	(0.013)	(0.013)	(0.013)
neighbors	0.135***	0.135***	0.114***	0.135***
	(0.019)	(0.019)	(0.018)	(0.019)
area	0.0001***	0.0001***	0.0001***	0.0001***
	(0.00001)	(0.00001)	(0.00001)	(0.00001)
terrain_ruggedness	-0.001***	-0.001***	-0.002***	-0.001***
	(0.0004)	(0.0004)	(0.0003)	(0.0004)
region_nankai	-0.039			
	(0.070)			
region_sanin		0.057		
		(0.046)		
region_sanyou			-0.074	
			(0.057)	
region_saikai				-0.127**
				(0.062)
Constant	-0.146	-0.146	0.017	-0.146
	(0.160)	(0.160)	(0.145)	(0.160)
Observations	8,442	8,442	8,442	8,442
R2	0.136	0.136	0.136	0.136
Adjusted R2	0.116	0.116	0.116	0.116
Residual Std. Error (df = 8244)	0.345	0.345	0.345	0.345
F Statistic (df = 197; 8244)	6.611***	6.611***	6.611***	6.611***
Note:		*p<0.1; **p<0.05; ***p<0.01		

Sixth, I reran the linear probability analysis with the same controls, with robust standard errors, and with two-way fixed effects, but with dyadic rather than monadic data, and with numerical (count) variables for the IVs and DV rather than dichotomous variables. The unit of analysis in this case is the province-dyad-year, including one observation for each pair of adjacent provinces for each year. The DV is a count of the number of battles taking place in a given province in a given calendar year (*prov1_battle*). The key IVs are a count of the number of battles taking place in an adjacent province in the prior calendar year (*prov2_battle*) and a count of the number of battles taking place in an adjacent province sharing a major road in the prior calendar year (*road_battle*). Table A12 presents the results. As is clear, the main results are unchanged and the size of the coefficient on *road_battle* in model 2 is substantively very similar to the main analysis (0.033 → 0.031), though it is only significant at the 0.05 level in this case. The coefficient on *road_battle* in model 3 is substantively much larger than in the main analysis (0.029 → 0.048).

Table A12: Linear Probability Analysis with Dyadic Data

	Dependent variable: prov1_battle		
	(1)	(2)	(3)
prov2_battle	0.036***	0.018*	
	(0.007)	(0.010)	
road_battle		0.031**	0.048***
		(0.013)	(0.009)
neighbors	0.324***	0.323***	0.323***
	(0.036)	(0.036)	(0.036)
area	0.0002***	0.0002***	0.0002***
	(0.00002)	(0.00002)	(0.00002)
prov1_terrain_ruggedness	-0.005***	-0.005***	-0.005***
	(0.001)	(0.001)	(0.001)
Constant	0.235	0.232	0.231
	(0.317)	(0.317)	(0.317)
Observations	33,782	33,782	33,782
R2	0.119	0.120	0.119
Adjusted R2	0.114	0.114	0.114
Residual Std. Error	1.078 (df = 33586)	1.078 (df = 33585)	1.078 (df = 33586)
F Statistic	23.334*** (df = 195; 33586)	23.271*** (df = 196; 33585)	23.361*** (df = 195; 33586)
Note:			*p<0.1; **p<0.05; ***p<0.01

Works Cited

Black, Nathan. 2013. "When Have Violent Conflicts Spread? Introducing a Data set of Substate Conflict Contagion." *Journal of Peace Research* 50 (6): 751-759.

Forsberg, Erika. 2014. "Diffusion in the Study of Civil Wars: A Cautionary Tale." *International Studies Review* 16 (2): 188-198.

JapanKnowledge Lib. 2022. Tokyo: NetAdvance, Inc. Available at: https://japanknowledge.com.