War Games:
North Korea's Reaction to US and South Korean Military Exercises

Vito D'Orazio

Since 1976, the militaries of the United States and South Korea have been holding routine joint military exercises (JMEs) for the purposes of military training and deterrence against North Korea. These exercises are frequently cited as a cause of tension on the peninsula, causing North Korea to escalate its conflictual rhetoric and behavior. I empirically assess this claim using new data on US-ROK JMEs and machine-coded event data collected by the Integrated Crisis Early Warning System. The findings show that North Korea does not systematically escalate its conflictual rhetoric or behavior during or near the occurrence of JMEs. The results hold for both low- and high-intensity exercises and for rhetoric that has the United States and South Korea as its target. **KEYWORDS:** joint military exercise, conflict, deterrence, escalation, event data, North Korea

SINCE THE END OF THE KOREAN WAR IN 1953, THE GOVERNMENTS OF the Republic of Korea (ROK, South Korea) and the United States have undertaken actions to deter another war on the Korean peninsula.[1] One such action is when their militaries execute joint military exercises (JMEs). In a classic security dilemma, many of the actions taken to promote long-term stability and security may actually cause short-term instability and risk conflict escalation (Jervis 1976). It is commonly argued that US-ROK JMEs incur such a trade-off (Chu 2006; Rabiroff 2010).

Ever since the inception of routine US-ROK JMEs in 1976, the Democratic People's Republic of Korea (DPRK, North Korea) has loathed these exercises, often placing their military on alert and calling them a dress rehearsal for war when they occur (Chu 2006; Farrell 2009; Oberdorfer 2001). While high tensions and occasional low-level military conflict are characteristic of the relationship between North and South Korea and between North Korea and the United States, JMEs are

often identified as a specific cause of such tensions. Implicit in this ar-gument is that the DPRK escalates its conflictual rhetoric and behavior in response to the JMEs.

I argue this is not the case; US-ROK JMEs do not trigger a system-atic escalation in conflictual rhetoric or behavior. The DPRK's rhetoric and behavior is routinely aggressive and militaristic, making any re-sponse to these exercises difficult to distinguish from normalcy. Fur-thermore, this is the case for *all* types of JMEs, whether they are of higher intensity and involve the physical movement of military assets or are of lower intensity where the training is carried out largely via com-puter simulations.

I conduct straightforward statistical tests to examine the relation-ship between US-ROK JMEs and North Korean activity more closely. To do so, I use new data on US and South Korean JMEs from 1998 through 2010 and high-frequency, machine-coded event data produced by the Integrated Crisis Early Warning System (ICEWS) project of the Defense Advanced Research Projects Agency (DARPA) (O'Brien 2010). To appropriately assess the hypotheses, the JMEs are coded based on their intensity, and the event data are aggregated into weekly counts of verbal conflict and material conflict from North Korea di-rected toward South Korea or the United States. The results from sev-eral negative binomial regressions generally support the hypothesis that these exercises are not sparking a systematic increase in conflictual rhetoric or behavior from the DPRK.

I begin with some background information on JMEs in general and US-ROK JMEs in particular. The hypotheses are then presented and as-sessed based on anecdotal evidence, descriptive evidence for a single year, and a systematic statistical evaluation. I conclude with policy implications for the future of JMEs held between the United States and South Korea.

Background on Joint Military Exercises

Joint military exercises, or joint war games, take place when the mili-taries from more than one state interact in such a way as to mutually en-hance their ability to carry out military operations. The United States and South Korea have been engaging in routine JMEs since 1976. These exercises couple an extensive practical training element with a strong strategic demonstration of support.

In the long term, US-ROK JMEs are held to signal a unified posi-tion between South Korea and the United States and to deter potential challengers to that position—particularly North Korea (Sharp 2011).

However, actions taken to enhance a state's security in the long term often cause instability and risk conflict escalations in the short term (Huth 1988; Huth and Russett 1984; Jervis 1976; Morgan 1983). Military exercises, particularly when held in a location where stability is as tenuous as that on the Korean peninsula, potentially run such risks. The increase in the number of US troops on the peninsula, the presence of US warships in nearby waters, and live-fire drills, among other characteristics of these exercises, may all be a cause of instability and an impetus for escalation by the DPRK.

During the Cold War, these exercises and the associated risks were seen as necessary to deter another war on the Korean peninsula. Team Spirit, a Cold War–era JME involving the United States and South Korea, grew to be the largest military exercise ever held between multiple states (Farrell 2009). By comparison, at its peak, Team Spirit was almost four times the size of Reforger, NATO's largest JME in Europe (Blackwill and Legro 1989). North Korea loathed Team Spirit, and its use as a bargaining chip by the United States during the negotiations of the landmark 1994 Agreed Framework is well documented (Gurtov 1996; Oberdorfer 2001).

With the conclusion of the Cold War and the diminished possibility of war in Korea, the scale of US-ROK JMEs has been drawn down, but the exercises continue. With Team Spirit's cancellation in 1994, Ulchi Focus Lens and Foal Eagle—two Cold War–era exercises that continue today—grew in significance. The Reception, Staging, Onward Movement, and Integration exercise (RSOI) was introduced in 1995 to make up for essential military training that was lost with the cancellation of Team Spirit (Chu 2006). In 2002, RSOI was combined with Foal Eagle, and now the two are held concurrently.[2] In 2008, Max Thunder, a JME involving the air forces of the ROK and the United States, was held for the first time and has been held every year since. In addition to these annual JMEs, others are conducted occasionally in reaction to North Korean aggression. One example is Invincible Spirit, a JME held in July 2010 in response to North Korea's sinking of the *Cheonan,* a South Korean warship (Garamone 2010; Sharp 2011).

During the timing of these exercises, it is common for the DPRK to verbally lash out against the United States and South Korea. In his illustrative study on Korean relations, Oberdorfer (2001, 273) recounts an anecdote where "a U.S. official who visited Pyongyang in 1993 said [Kim Il Sung's] voice quivered and his hands shook with anger when he discussed Team Spirit in a conversation with Representative Gary Ackerman." In talking about the RSOI/Foal Eagle exercise in 2002, an

Agence France-Presse (2002) article states, "North Korea has said the drills are 'all-round war exercises aimed to put the plan for preemptive strikes' at the north into practice and 'very dangerous war gambles to seize the chance to provoke a nuclear war.'" In response to another JME, an Associated Press article from 2008 quotes the North Korean govern-ment's state-run news agency as saying, "[North Korea] will sternly take self-defensive steps to defend [the] peace and stability of the Ko-rean peninsula" (Kim 2008).[3] And in a more recent statement to the US Senate Armed Services Committee, General Walter Sharp noted that "North Korea threatened to attack ROK and US forces participating in the Key Resolve exercise" (Sharp 2011).

In addition to aggressive rhetoric, it is routine for the DPRK to be-have in militaristic ways as well. Examples include placing its military on alert during the time of US-ROK JMEs and/or countering with some sort of military exercises of its own.[4]

Reports such as these have led one military analyst to state that "military exercises have, in fact, adversely strained diplomatic relations with North Korea" (Chu 2006, 1). Many others have come to similar conclusions and question whether US-ROK JMEs are worth the risks associated with them (Rabiroff 2010). Implicit in these arguments is not just that North Korea responds aggressively to US-ROK JMEs but that the aggressive rhetoric and behavior are unique and systematic re-sponses to the exercises. I argue in the following discussion that this is not the case. I argue in the following discussion that this is not the case.

US-ROK JMEs and Conflict Behavior: Hypotheses

While it may be true that these JMEs elicit an aggressive response, I argue that this response is neither unique nor systematic to all US-ROK war games. That the DPRK response is not unique means that the DPRK will respond in a similarly aggressive fashion to many types of activities involving the United States and South Korea. While there are some interactions that do trigger a uniquely aggressive response, two of which are highlighted below, US-ROK JMEs are not one of them. That the DPRK response is not systematic means that North Korea does not respond in a common way to all JMEs. Effectively, the DPRK response to US-ROK JMEs is routine and thus statistically indistinguishable from normalcy.

Part of the reason North Korea's response is routine is due to the in-ternational norm of advanced notification given to states that may deem the military exercise a threat. Arguably, North Korea knows in advance

that the exercise is just an exercise and has ample time to prepare a routine response.[5] The exercises are often viewed as escalating tensions beyond what is normal probably because of increased media attention at the time of the exercises. When the media report that North Korea has called the exercise a "dress rehearsal for war," the JME appears as a destabilizing provocation. However, if the mainstream media reported regularly on the statements of the DPRK, it would become clear that similarly flagrant statements are commonplace.

> *Hypothesis 1:* US-ROK JMEs are not associated with increases in conflictual behavior by North Korea.

While US-ROK JMEs in general might not cause escalation, not all JMEs are the same. In particular, some JMEs are more intense than others and therefore might be more threatening. Often, these are the exercises that attract more media attention and may evoke a larger response from the DPRK. Among the types of JMEs that increase intensity are live-fire drills, field training exercises, exercises involving an aircraft carrier, exercises involving a large number of troops, and reactionary exercises. In general, if any JME should be expected to be associated with increases in conflictual rhetoric and behavior among either of the dyads of interest—that is, between North Korea and South Korea or North Korea and the United States—it should be the ones of higher intensity.

There is evidence that the North Korean government takes notice of these attributes of the exercises. For example, the "White Paper of the National Reunification Institute," a transcript of which has been made available by the British Broadcasting Corporation, states that "the Key Resolve and the Foal Eagle joint military exercises being staged this time also greatly exceed those staged in the past in terms of duration and scale" (Section I). The report goes on to describe increases in the duration of the exercise as well as various increases in the magnitude of the exercise. More generally, however, the report is describing increases in the intensity of the exercise and claiming that such increased intensity indicates that the United States and South Korea are planning an attack (BBC 2011).

In a report for the United States military, John Chu argues along the lines that North Korea responds to increases in the intensity of US-ROK JMEs. He states that "the only option to remove key obstacles [to unification], short of complete withdrawal and further realignment or reduction of U.S. forces, is to make changes to the military exercises. What may appear as minor changes to the United States could mean renewed

trust and the possibility for cooperation and dialogue to North Korea"
(Chu 2006, 19). In his qualitative military report, Chu essentially con-
cludes that the United States should decrease the intensity of its JMEs
with South Korea. His analysis, however, does not systematically com-
pare types of JMEs.

> *Hypothesis 2A:* US-ROK JMEs of higher intensity are associated
> with more conflicual behavior by North Korea than US-ROK
> JMEs of lower intensity.

However, DPRK responses may not change depending on the in-
tensity of the exercise because of the fact that the DPRK has a high
baseline level of conflictual rhetoric and behavior to begin with. The
magnitude of the response might be larger, as Chu (2006) presumes, but
not to the degree that the response is systematically distinguishable
from normal behavior. Therefore, the relationship between high-inten-
sity JMEs and all other time periods may be the same as the expectation
for all JMEs.

> *Hypothesis 2B:* US-ROK JMEs of higher intensity are not associ-
> ated with increases in conflictual behavior by North Korea.

Data

The research design I use here falls under the category of a quantitative
case study. The approach is consistent with event data research on dis-
putes such as the Balkans conflicts and the Israel-Palestinian conflicts
(Pevehouse and Goldstein 1999; Schrodt and Gerner 2000; Shellman
2006). It is also consistent with at least one other event data paper on
North Korea (Yoon 2011).

For this article, I use new data collected on US-ROK JMEs from
1998 through 2010. I measure the rhetoric and actions of North Korea
using event data that are produced by the DARPA-funded ICEWS proj-
ect and that represent a new standard in high-frequency, machine-coded
event data (O'Brien, 2010). The event data are systematically aggre-
gated into weekly counts of verbal conflict and material conflict di-
rected from North Korea toward South Korea or the United States.

The ICEWS project has coded event data for the twenty-nine states
in the US Pacific Command's Area of Responsibility. To code this data,
ICEWS utilizes a collection of 6.5 million news documents from sev-
enty-five international news sources of this time period. These docu-

ments are processed by the ICEWS' adaptation of TABARI, natural language processing software engineered to extract semantic content from natural language in a who-did-what-to-whom-and-when format (Gerner et al. 1994; Schrodt and Gerner 1994). In other words, ICEWS uses TABARI as a machine coder; users input news documents, and it outputs raw event data.

Raw event data consist of four variables coded for each event (or observation): the date the event took place, the action (what was done), the source (who did it), and the target (who it was done to). ICEWS' version of TABARI extracts each of these variables from a news document and then codes them according to actor dictionaries developed by the ICEWS project and an action dictionary provided by CAMEO. These dictionaries are essentially a mapping from the actors or actions extracted by TABARI to a simplified coding scheme (Schrodt 2009).

An example is helpful to better understand how the data are collected. Let us begin with a news document that reads, "Thailand severed relations with Laos today." ICEWS' TABARI would extract a date (whatever date the news story was reported), a source actor (Thailand), a target actor (Laos), and an action (severed relations). The actor dictionary is then used to code each of the actors according to its country code—in this case, THA and LAO. The action is also coded according to the CAMEO code under which it falls. In this case, "severed relations" would be coded as CAMEO code 161, which is the code for "reduced or broke diplomatic relations" (Schrodt and Yilmaz 2007).

This process is undertaken for the entire collection of news stories, providing us with raw event data; every row in the data is a unique event for which we have four variables coded: date, source, target, action. Raw event data are typically not useful for statistical analysis, however. They must be aggregated in some meaningful fashion. To do so, there are three points of aggregation to consider: actor aggregations, action aggregations, and temporal aggregations. Each of these is discussed in turn as it relates to the analysis presented.

Actor aggregations concern the source and target. CAMEO provides information about the source and target on the national level as well as on various subnational levels. In the preceding example, the actors were clearly on the national level: Thailand is the source and Laos the target. However, news stories are often not so clear. For example, let us say the news document reads, "Abhisit Vejjajiva severed relations with his counterpart in Laos."[6] TABARI would extract "Abhisit Vejjajiva" as the source actor, not Thailand. The actor dictionary would recognize that Abhisit Vejjajiva is the prime minister of Thailand, and the

source actors would be coded as THAGOV, signifying Thailand as the national actor and a member of government as the subnational actor. This is an example of two-tiered actor coding. When possible, ICEWS' TABARI can code three-tiered actor coding. For our purposes, I chose to aggregate actors by using all events where (1) the source at the national level is North Korea and the target at the national level is United States, or (2) the source at the national level is North Korea and the target at the national level is South Korea. I dismiss second- and third-tier actor data.

Action aggregations involve managing the different types of events that could occur between actors. CAMEO's actions dictionary codes for twenty categories of actions, ranging from "Make Public Statement" to "Engage in Unconventional Mass Violence" (Schrodt and Yilmaz 2007). There are many ways to handle the action aggregation when so many categories exist. For example, twenty categories could result in twenty variables, where each variable is a count of the number of times the actors engaged in that category of action. Or, the twenty variables could be placed on a scale such as the Goldstein Scale (Goldstein 1992). Scaling data, however, requires us to assume that the categories exist on a single continuum and at fixed intervals. These are assumptions that I want to avoid because there is no objective method for stating that "exhibit force posture" is x units less than "reduce relations," which is w units less than "coerce," and so forth.

The approach taken stems from the Duval and Thompson (1980) method of collapsing each event into one of four classes: verbal conflict, material conflict, verbal cooperation, and material cooperation. Here, only verbal and material conflict are used because we are interested in explaining conflictual, not cooperative, behavior. The action aggregation method used is a count of the number of times the actors engaged in verbal conflict or material conflict for some time period. Table 1 lists the component categories for each of the two classes.

Table 1 Component Categories of Verbal and Material Conflict

Verbal Conflict	Material Conflict
Demand	Exhibit force posture
Disapprove	Reduce relations
Reject	Coerce
Threaten	Assault
Protest	Fight
	Use conventional mass violence

The time period is the final point of aggregation for transforming raw event data into usable data for empirical analysis. Data may be aggregated at nearly any unit of time: day, week, month, year, etc. The temporal aggregation must be appropriate for the question asked because different temporal aggregations can lead to different results (Alt, King, and Signorino 2001; Shellman 2004). In most quantitative studies in international relations, data are aggregated annually. The results of these studies are generalized trends across years. Event data, however, are fine-grained to the point where we can utilize much lower levels of temporal aggregation, and therefore those data can be used to discuss generalized trends by week, month, and so forth.

The unit of time used here is the week, which is perhaps the most natural unit of time for event data research because all weeks (with the exception of the very last week in each year) follow the same seven-day format and include exactly five weekdays (when news reporting is heavier) and two weekend days (when news reporting is lighter). For our purposes, the week is also the most appropriate since JMEs tend to be one or two weeks in length. If we had used a monthly aggregation, we would have had to code the JME as having taken place for the entirety of a month and therefore associate the JME with many events that did not take place during the actual time of the JME. However, if we had used a daily aggregation, we would never have miscoded any events as having occurred during the JME, but that data would have been very sparse. With past event data collection projects, even weekly aggregations would probably be too sparse to analyze. However, with ICEWS' project, this is no longer a problem because of the massive amounts of data collected. In sum, the event data are aggregated into four variables: weekly counts of verbal conflict and material conflict that take place between either the DPRK-US or DPRK-ROK dyad.

To collect data on US-ROK JMEs, research, statements, and general searches on the Web were used to create a list of all the names of the various exercises. A variety of news sources were then searched via LexisNexis for information on these events. Sources include the *New York Times, Korea Times, Korea Herald,* Xinhua General News Service, Associated Press, Agence France-Presse, and the British Broadcasting Corporation. The data collected using these sources were followed by searches on the Web to improve their validity.

Some have argued that the intensity of the military exercise may be a cause for escalatory behavior on the part of North Korea. Some evidence has been presented that the DPRK does take notice of the intensity of the exercise and that some US-ROK JMEs are more intense than

others. To assess this hypothesis more appropriately, field-training exercises and reactionary exercises are coded as more visible than others. A field-training exercise involves the physical movement of military assets; this is in contrast to a command-post exercise in which troop movements are simulated by computer. Reactionary JMEs are those that are not held on a regular basis, but rather have occurred as a response to some action taken by North Korea. For the purposes of this study, Foal Eagle/RSOI and the set of nonroutine US-ROK JMEs that followed the sinking of the *Cheonan* have been coded as high-intensity.

Descriptive Evidence

Before conducting any statistical analysis, it will be fruitful to describe the data more thoroughly and to view some of it graphically. To visualize which JMEs are occurring and when, a graphical representation of all JMEs used in the statistical analysis is provided in Figure 1. As we can see, Ulchi Focus Lens has occurred every year from 1998 through 2010. Foal Eagle has been held every year with the exception of 2001, when it was canceled in the aftermath of the September 11 attacks on the United States. When Foal Eagle was resumed, it was combined with

Figure 1   US-ROK JMEs, 1998–2010
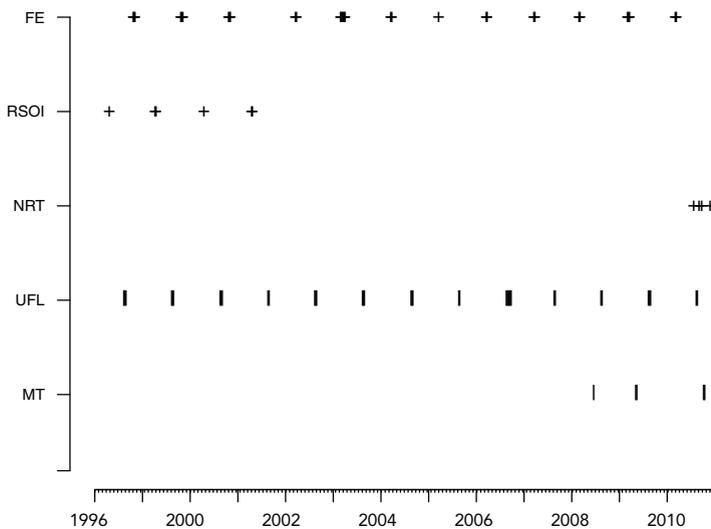
Note: + denotes higher visibility; | denotes lower visibility.

the other JME of higher intensity, RSOI. The only nonroutine JMEs that have been coded all took place in 2010 and include the Invincible Spirit exercise as well as the antisubmarine exercises. Max Thunder, begun in 2008, also occurred in 2009 and 2010.

Figure 2 shows the event data for the year 2008. The plotted solid line in each graph corresponds to the event counts, aggregated as described previously. The vertical, dashed lines represent the occurrence of a US-ROK JME for that week in 2008. The three exercises plotted are, in order from left to right, Foal Eagle/RSOI, Max Thunder, and Ulchi Focus Lens. In the graph in the upper left of the figure, both Foal Eagle/RSOI and Ulchi Focus Lens appear to be associated with increases in verbal conflict directed toward the United States. Max Thunder does not appear to have any association. Furthermore, there appears to be no recognizable association between JMEs and the conduct of material conflict directed toward the United States, nor any association between either verbal conflict or material conflict directed toward South Korea.

Figure 2 Event Data and US-ROK JMEs, 2008
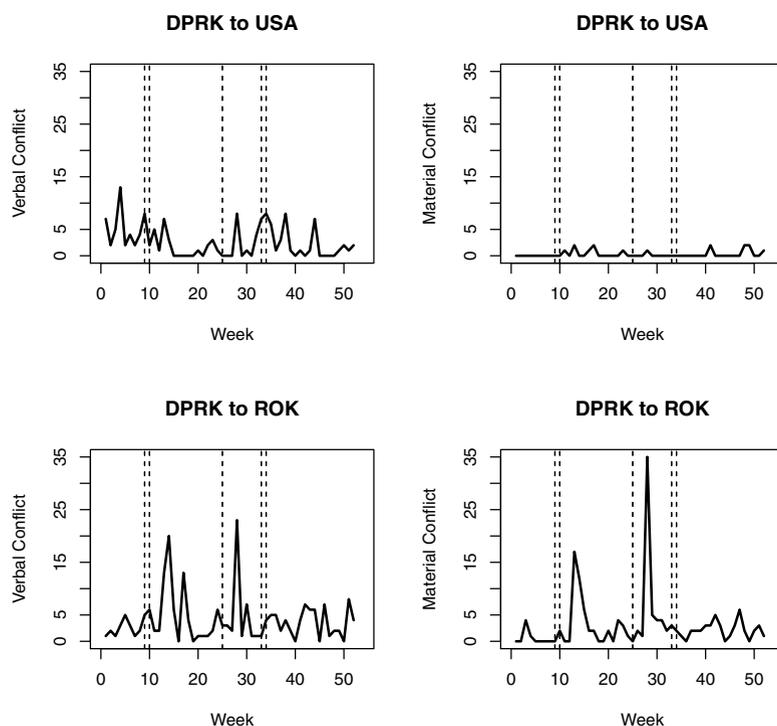
In addition to providing graphical evidence that generally supports the hypotheses for the year 2008, Figure 2 also serves as a validity check for the event data used to systematically evaluate these hypotheses. I have earlier argued that JMEs do not trigger unique behavioral changes from North Korea but that other types of events do trigger these changes. If the data collected are to be used to adequately assess the hypotheses, then we should first check to make sure that the data do, in fact, capture those episodes of unique behavioral changes from North Korea.

Counts of verbal conflict and material conflict from the DPRK to the ROK are presented in the lower two graphs in Figure 2. In each of these graphs there are two sets of peaks that appear unique from the rest of the image and represent increased conflictual behavior on the part of North Korea. The first of these behavioral changes takes place in late March 2008. During this period, North Korea had been claiming that South Korea had violated its maritime border in the Yellow Sea multiple times. North Korea also conducted several missile tests during this time that drew international attention (Sang-hun 2008b). The second set of peaks corresponds to a mid-July killing of a South Korean tourist in the North (Sang-hun 2008a). Following the killing, the DPRK refused to allow ROK personnel into the country to investigate, sparking a major disruption in relations between the two states. These distinguishable increases in tensions are captured quite clearly by the high-quality event data produced by the ICEWS project.

Empirical Analysis

To adequately assess the hypotheses with these data, descriptive evidence of just one year is not sufficient. To systematically analyze all data, twelve negative binomial regressions are estimated, four for each hypothesis. The four regressions for each hypothesis correspond to the four possible dependent variables: verbal conflict toward the United States, material conflict toward the United States, verbal conflict toward the ROK, and material conflict toward the ROK. The models assessing Hypothesis 1 utilize three binary variables coding the JME during the week(s) in which the exercise occurred, the two weeks after the exercise, and the two weeks prior to the exercise. Hypotheses 2A and 2B are assessed using disaggregations of the JME variables based on exercise intensity.

All four dependent variables are count variables. When estimating a model where the dependent variable is a count, the standards approaches are to use either a Poisson or a negative binomial regression.

Because the Poisson assumes that the mean is equal to the variance, the decision hinges on whether or not the dependent variable has a mean that is close enough to the variance to not be overdispersed. A simple test for overdispersion, as described in Cameron and Trivedi (1998, 77) is to see if the sample variance of the count variable is greater than twice its mean. The results of this test demonstrate that this is the case for each of the four dependent variables, meaning that overdispersion is present and a negative binomial regression is more appropriate.

Each of the four models assessing Hypothesis 1 contains four explanatory variables to test for associations at different points in time and to control for general time trends: *jme*, *lag_jme*, *lead_jme*, and *log_year*. The three JME variables are dichotomous and indicate either the presence of a US-ROK JME for that week (*jme*), the presence of a US-ROK JME in one of the following two weeks (*lag_jme*), or the presence of a US-ROK JME in one of the previous two weeks (*lead_jme*). It is important to include the *lag_jme* variable because of the fact that the United States and South Korea, in accordance with international standards, always provide advance notice to North Korea of a pending JME. The *lead_jme* variable is intended to capture a reaction by North Korea that may take several days to come to fruition or to capture an increase in tensions that may take some time to deescalate. For example, both increases in hostilities described in Figure 2 by the two sets of peaks in the graphs take longer than one week to return to normal relations. If JMEs are causing similar shocks, then a similarly slow decay in hostilities is to be expected. The *log_year* variable controls for linear trends in time.

The results generally support the first hypothesis (see Table 2). In the four models assessing Hypothesis 1, only two of the twelve coefficients are statistically significant and positive, meaning the presence of a US-ROK JME is not associated with changes in DPRK activity in the ten other categories. The two coefficients that are statistically significant and positive are material conflict directed toward the United States in the time period of the JME and material conflict directed toward South Korea in the two-week period preceding the JME. Of the remaining ten coefficients that are not statistically significant, five of them are negative. While the standard error of these estimates is too large to report with confidence that the association is negative, the mixture of positive and negative coefficients tells us that even if we lower the acceptable *p*-values, we would still not have overwhelming evidence that US-ROK JMEs increased conflictual behavior by North Korea.

Hypotheses 2A and 2B suggest that disaggregating JMEs into those of higher and lower intensity might help distinguish the exercises that

Table 2 Assessing Hypothesis 1

	North Korea to United States		North Korea to South Korea	
	Material Conflict	Verbal Conflict	Material Conflict	Verbal Conflict
jme	0.466*	0.202	0.229	–0.008
	(0.247)	(0.131)	(0.182)	(0.125)
lag_jme	0.237	–0.015	0.538*	0.057
	(0.265)	(0.141)	(0.188)	(0.131)
lead_jme	0.383	–0.134	–0.105	–0.056
	(0.263)	(0.142)	(0.200)	(0.133)
log_year	201*	–47.9*	242*	62.3*
	(44.9)	(23.2)	(32.9)	(21.9)
intercept	–1527*	363*	–1837*	–473*
	(341)	(177)	(250)	(167)

Notes: * indicates p-value < 0.1; standard error in parentheses.

are causing more of a response and those that are not. This is done by splitting the JME variables into two categories, *h_intense* and *l_intense*.[7] Each of these variables, in addition to *log_year*, is included in the four models assessing Hypothesis 2B. These results, shown in Table 3, are generally consistent with those from the first regression table and with the expectation that even US-ROK JMEs of high intensity are not systematically associated with increased tensions. However, the two coefficients that are statistically significant are statistically significant only in the case of high-intensity JMEs. But, if we look at the signs of the other *h_intense* coefficients, they too are consistent with the regression results above: five negative categories and five positive categories. Of the twelve coefficients on low-intensity JMEs, none are statistically significant, four are negative, and eight are positive. Overall, there is roughly an equivalent amount of support for Hypothesis 2B as there is for Hypothesis 1. The expectation that neither *h_intense* nor *l_intense* JMEs will heighten tensions is generally supported.

The expectation from Hypothesis 2A is that *h_intense* JMEs should be associated with higher levels of tensions than *l_intense* JMEs. Since the results from Table 3 are interpreted in reference to the omitted category—that is, the periods of time that are not temporally near US-ROK JMEs—it is difficult to appropriately assess this hypothesis using those results. As a first take, however, the categories driving the statistical significance in Table 2 are *h_intense* categories in Table 3. Furthermore, where the coefficients are positive, the magnitude is generally higher for *h_intense* categories than for *l_intense* ones. While this may

Table 3 Assessing Hypothesis 2B

	North Korea to United States		North Korea to South Korea	
	Material Conflict	Verbal Conflict	Material Conflict	Verbal Conflict
h_intense jme	0.769*	0.223	0.315	–0.046
	(0.329)	(0.179)	(0.248)	(0.172)
h_intense lag_jme	–0.023	0.125	0.731*	0.165
	(0.354)	(0.182)	(0.242)	(0.17)
h_intense lead_jme	0.472	–0.289	–0.019	–0.081
	(0.342)	(0.189)	(0.261)	(0.176)
l_intense jme	0.012	0.185	0.138	0.032
	(0.341)	(0.178)	(0.247)	(0.168)
l_intense lag_jme	0.055	–0.212	0.224	–0.086
	(0.371)	(0.208)	(0.278)	(0.193)
l_intense lead_jme	0.031	0.029	–0.133	–0.015
	(0.378)	(0.204)	(0.287)	(0.192)
log_year	197*	–50.0*	242*	63.1*
	(44.7)	(23.2)	(32.9)	(22.0)
intercept	–1500*	379*	–1840*	–479*
	(340)	(177)	(250)	(167)

Notes: * indicates p-value < 0.1; standard error in parenthesis

imply some support, it is not sufficient. To get a more direct comparison between the two types of exercises, the omitted category is switched from *no_jme* to *l_intense*; *h_intense and l_intense* are revised to a binary coding where the two weeks before the JME, the weeks of the JME, and the two weeks after the JME are all coded as 1.

The results from the final set of models are reported in Table 4. In these models, *h_intense* is interpreted with respect to the omitted category, *l_intense*. The coefficient on material conflict for the DPRK-ROK dyad is the only coefficient that is statistically significant and positive. This suggests that, for material conflict for DPRK-ROK, the response to high-intensity JMEs is systematically different from the response to low-intensity JMEs—even if the response is not systematically different from normal DPRK behavior. Of the remaining three coefficients, two are positive and one is negative. Given these results, although there is *some* evidence that high-intensity JMEs are associated with escalations of greater magnitude than low-intensity ones, we cannot conclude that *h_intense* JMEs are systematically different from *l_intense* JMEs.

The statistical tests conducted here provide evidence that we are not more likely to observe behavioral changes from the DPRK during the timing of US-ROK JMEs. This could mean two things. One, US-ROK

Table 4 Assessing Hypothesis 2A

	North Korea to United States		North Korea to South Korea	
	Material Conflict	Verbal Conflict	Material Conflict	Verbal Conflict
h_intense	0.168	–0.030	0.350*	0.037
	(0.284)	(0.153)	(0.211)	(0.144)
no_jme	–0.322	–0.101	–0.072	0.014
	(0.235)	(0.124)	(0.174)	(0.117)
log_year	201*	–46.4*	245*	62.8*
	(44.6)	(23.2)	(32.8)	(21.9)
intercept	–1531*	355*	–1860*	–476*
	(339)	(177)	(249)	(166)

Notes: * indicates p-value < 0.1; standard error in parentheses.

JMEs do not trigger a reaction from North Korea at all; these exercises come and go and the DPRK ignores them. Given the ample anecdotal evidence that North Korea frequently mentions US-ROK JMEs, and that Team Spirit was an influential tool for bargaining in the 1994 Agreed Framework, this case is extremely unlikely. More likely to be true is the argument that North Korea does respond with low-level aggression to US-ROK JMEs, but that it also responds aggressively to a whole host of other events—so much so that in these data the presence of a JME is not associated with any unique increases in tensions.

Discussion and Conclusion

North Korea has posed a strategic challenge for the governments of the United States and South Korea for well over half a century. The primary dilemma for these governments has been and remains how to deter the DPRK while also not taking actions that risk destabilizing the situation and provoking escalations toward conflict. To this end, one sustained policy has been to hold joint military exercises.

During the Cold War, these exercises were viewed as a necessary signal of strength for the purposes of deterring war. Today, however, the chances of war on the Korean peninsula are slim, and if war did break out, the DPRK would be easily outmatched. Because of this, US-ROK JMEs have come under criticism. As during the Cold War era, they are argued to elicit an aggressive North Korean response and risk conflict escalation. However, due to the diminished possibility of war in the current era, their signal of deterrence is said to be not as necessary as it

once was. An implicit point in this argument is that the exercises actually do escalate conflictual rhetoric and/or behavior by North Korea.

I have demonstrated empirically that this point is largely invalid. The DPRK's response to US and South Korean joint exercises is not unique and not systematically different from regular DPRK activity. While some evidence exists that high-intensity exercises are associated with a response that is greater in magnitude and different from low-intensity JMEs, the evidence is inconclusive and not systematic across types of DPRK behavior.

For the governments of these states, the findings in this study could be useful for formulating policy around JMEs. In particular, while canceling or even reducing the intensity of such exercises may be an appropriate gesture of goodwill, and while it may be necessary to enact certain policies, these governments should not expect such an action to fundamentally change the rhetoric or behavior of North Korea. Tensions will still flare, and aggressive statements will still occur routinely.

Future quantitative research on North Korean activity has several promising avenues, especially given the recent advances in event data collection methods. From the perspective of domestic politics, North Korea's behavior can be assessed based on the political party in power in the United States or South Korea. With the rise to power of Kim Jong Un, the DPRK's rhetoric and behavior across regimes and during the transition could be examined. Future research involving JMEs could look at reactionary exercises, which may have a distinguishable effect from other JMEs because the exercise itself may be seen as conflict-escalating. The number of reactionary exercises is too few to conduct such a test solely on North Korea, but such tests could be done with a cross-national research design.

Vito D'Orazio is a PhD candidate in the Department of Political Science at Pennsylvania State University. He is a predoctoral fellow at Penn State's Quantitative Social Science Initiative where he researches applications of machine learning methods to international relations. His dissertation examines dyadic and systemic security cooperation using item response theory and social network analysis. Other research interests include using sequence analysis to predict civil conflict and exploring automated methods of data collection.

Notes

I presented previous versions of this essay at the annual meeting of the International Studies Association, Montreal, March 16–19, 2011, and at the annual

meeting of the American Political Science Association, Seattle, September 1–4, 2011. I would to thank David Carter, Stephan Haggard, Margaret Kosal, Matthew Kroenig, Glenn Palmer, and Phil Schrodt for their comments and assistance with this project. This project was funded in part by a contract from the Defense Advanced Research Projects Agency under the Integrated Crisis Early Warning System (ICEWS) program (Prime Contract #FA8650-07-C-7749: Lockheed-Martin Advanced Technology Laboratories). The views, opinions, and findings contained in this paper are mine alone and should not be interpreted as representing the official views or policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the Department of Defense, nor those of Lockheed-Martin.

1. I refer to the Democratic People's Republic of Korea as either North Korea or the DPRK. I refer to the Republic of Korea as either South Korea or the ROK.

2. By 2008, Foal Eagle/RSOI had been renamed Key Resolve, and Ulchi Focus Lens had been renamed Freedom Guardian.

3. The news stories that contain evidence of North Korea lashing out against the United States and South Korea over these JMEs easily number in the hundreds if not thousands.

4. Normally, placing one's military on alert would be coded as a militarized interstate dispute (MID) according to the Correlates of War (COW) project. However, this activity is so routine that COW does not treat this action by the DPRK as a MID.

5. This norm was laid out for Europe in the Helsinki Accords of 1975 and expanded in the Stockholm Accords of 1986. It has since grown into something of an internationally accepted protocol.

6. Abhisit Vejjajiva was the prime minister of Thailand from 2008 to 2011.

7. Reactionary JMEs are coded as *h_intense* exercises, although they may not have all the same attributes of the other *h_intense* JMEs. However, there are only four reactionary exercises, and their inclusion in either category or their exclusion in the models altogether does not alter the interpretation of the results.

References

Agence France-Presse (AFP). 2002. "US and South Korea Kick Off Biggest Military Drills." March 21.

Alt, James, Gary King, and Curt S. Signorino. 2001. "Aggregating Among Binary, Count and Duration Models: Estimating the Same Quantities from Different Levels of Data." *Political Analysis* 9, 1: 21–44.

BBC. 2011. "White Paper of the National Reunification Institute—We Condemn the United States and the Puppet Warmongers' War Exercise Manoeuvre of Northward Aggression." *BBC Monitoring Asia Pacific*, March 12.

Blackwill, Robert D., and Jeffrey W. Legro. 1989. "Constraining Ground Force Exercises of NATO and the Warsaw Pact." *International Security* 14, 3: 68–98.

Cameron, Colin A., and Pravin K. Trivedi. 1998. *Regression Analysis of Count Data*. New York: Cambridge University Press.

Chu, John S. 2006. "Military Exercises in Korea: A Provocation or a Deterrent to War?" Newport, RI: Naval War College.

Duvall, Robert D., and William R. Thompson. 1980. "Reconsidering the Aggregate Relationship Between Size, Economic Development, and Some Types of Foreign Policy Behavior." *American Journal of Political Science* 24, 3: 511–525.

Farrell, John F. 2009. "Team Spirit: A Case Study on the Value of Military Exercises as a Show of Force in the Aftermath of Combat Operations." *Air and Space Power Journal* 23, 3: 95–106.

Garamone, Jim. 2010. "U.S.-Korean Defense Leaders Announce Exercise Invincible Spirit." American Forces Press Service, July 20. Available at www.defense.gov/news/newsarticle.aspx?id=60074 (accessed January 27, 2012).

Gerner, Deborah J., Philip A. Schrodt, Ronald A. Francisco, and Judith L. Weddle. 1994. "The Machine Coding of Events from Regional and International Sources." *International Studies Quarterly* 38: 91–119.

Goldstein, Joshua S. 1992. "A Conflict-Cooperation Scale for WEIS Events Data." *Journal of Conflict Resolution* 36, 2: 369–385.

Gurtov, Mel. 1996. "South Korea's Foreign Policy and Future Security." *Pacific Affairs* 69, 1: 8–31.

Huth, Paul. 1988. "Extended Deterrence and the Outbreak of War." *American Political Science Review* 82, 2: 423–443.

Huth, Paul, and Bruce Russett. 1984. "What Makes Deterrence Work: Cases from 1900–1980." *World Politics* 36, 4: 496–526.

Jervis, Robert. 1976. *Perception and Misperception in International Politics*. Princeton: Princeton University Press.

Kim, Kwang-Tae. 2008. "SKorea, US Begin Annual Military Maneuvers Amid NKorean Protests." Associated Press, March 3.

Morgan, Patrick M. 1983. *Deterrence: A Conceptual Analysis.* 2nd ed. Beverly Hills: Sage.

North-South Joint Declaration. 2000. Signed June 15. Available at www.usip.org/files/file/resources/collections/peace_agreements/n_skorea06152000.pdf (accessed January 27, 2012).

Oberdorfer, Don. 2001. *The Two Koreas: A Contemporary History*. 2nd ed. New York: Basic Books.

O'Brien, Sean. 2010. "Crisis Early Warning and Decision Support: Contemporary Approaches and Thoughts on Future Research." *International Studies Review* 12, 1: 87–104.

Pevehouse, Jon C., and Joshua S. Goldstein. 1999. "Serbian Compliance or Defiance in Kosovo? Statistical Analysis and Real-Time Predictions." *Journal of Conflict Resolution* 43, 4: 538–546.

Rabiroff, Jon. 2010. "Experts Assess Merit, Risks of U.S.–South Korea Exercise." *Stars and Stripes,* July 23.

Sang-hun, Choe. 2008a. "North Korea Faults South in Death of Tourist and Rejects Talks Offer." *New York Times–Asia Pacific.* July 13. Available at www.nytimes.com/2008/07/13/world/asia/13korea.html (accessed January 27, 2012).

———. 2008b. "South Korea Plays Down Missile Test by the North." *New York Times–Asia Pacific.* March 29. Available at www.nytimes.com/2008/03/29 /world/asia/29korea.html (accessed January 27, 2012).

Schrodt, Philip A. 2009. TABARI: Textual Analysis by Augmented Replacement Instructions. Available at http://web.ku.edu/~keds/tabari.dir/tabari .manual.0.7.3b3.pdf (accessed January 27, 2012).

Schrodt, Philip A., and Deborah J.Gerner. 1994. "Validity Assessment of a Machine-Coded Event Data Set for the Middle East, 1982–1992." *American Journal of Political Science* 38: 825–854.

———. 2000. "Cluster-Based Early Warning Indicators for Political Change in the Contemporary Levant." *American Political Science Review* 94, 4: 803–817.

Schrodt, Philip A., and Omur Yilmaz. 2007. "Conflict and Mediation Event Observations (CAMEO) Codebook." Available at http://web.ku.edu/~keds /cameo.dir/CAMEO.CDB.09b5.pdf (accessed January 27, 2012).

Sharp, Walter L. 2011. Statement Before the Senate Armed Services Committee, April 12. Available at http://armed-services.senate.gov/statemnt/2011 /04%20April/Sharp%2004-12-11.pdf (accessed February 10, 2012).

Shellman, Stephen. 2004. "Time Series Intervals and Statistical Inference: The Effects of Temporal Aggregation on Event Data Analysis." *Security Studies* 12, 1: 97–104.

———. 2006. "Process Matters: Conflict and Cooperation in Sequential Government-Dissident Interactions." *Security Studies* 15, 4: 563–599.

Yoon, Jong-Han. 2011. "The Effect of US Foreign Policy on the Relationship Between South and North Korea: Time Series Analysis of the Post–Cold War Era." *Journal of East Asian Studies* 11, 2: 255–287.