The academic literature has explored many possible tangible and intangible factors affecting the final, and essentially political, outcomes of interstate wars. Many of the hypothesized mechanisms connecting these factors to war outcomes are based on theories of military effectiveness (Biddle and Long 2004, Reiter and Stam 1998). Yet, these mechanisms have not been tested because the quantitative literature focuses almost exclusively on categorical outcomes that reflect political settlements (Stam 1999, Slantchev 2004). Testing these mechanisms is key to adjudicating between different theories of war outcomes and to understanding military effectiveness more generally. While a state's ability to achieve favorable battlefield outcomes has a profound effect on its ability to protect its citizens from external threats and secure its interests abroad, war outcomes do not always reflect superior effectiveness on the battlefield. Disentangling war outcomes and military effectiveness can shed light on the relationship between the two, while also allowing us to develop and test theories focused specifically on the military dimension. Military effectiveness drives the relative costs of war and is in turn affected by material and political factors within the control of states. Understanding military effectiveness can help us better explain why some political victories in war come at such great cost, while others come with relative ease. Yet, our understanding of the determinants of battlefield effectiveness is limited because of existing data's temporal bounds, spatial scope, and selection bias.

In this paper, we introduce and utilize a new dataset that remedies these problems by providing casualty Loss Exchange Ratios for combatant states involved in multilateral wars between 1816 and 1990. These battle-level data provide an alternative to the widely used, but problematic, HERO/CDB-90 data set on battle characteristics. Our purpose in this paper is to describe this new resource and demonstrate one application of it so that other scholars may bring it to bear on important theoretical questions about military effectiveness in interstate war.

The Loss-Exchange Ratio (LER) for a military engagement is the ratio of casualties a combatant suffers to the casualties it inflicts on the enemy. It has been used as a standard measure of battlefield effectiveness in the defense community, but not by academics studying military effectiveness (Biddle 2004: 22). One reason for this disconnect between policy analysts and academic researchers is that although it is simple to calculate LERs for bilateral conflicts, it is nearly impossible to do so for multilateral conflicts given the format of most existing casualty datasets. These datasets provide casualty totals for each combatant state in a given war, but provide no information on who inflicted which casualties on whom. The few battle-level datasets that provide directed casualty information have temporally- or spatially-limited domains or suffer from non-random selection bias.

The Loss-Exchange Ratio Database (LERD) uses secondary historical sources to provide battle-level LERs for these multilateral wars, resulting in a data set of battles in multilateral wars from 1816 to 1990 that has 615 more battles than the HERO/CDB-90 dataset as cleaned and corrected by Biddle and Long (2004).

War Outcomes and Military Effectiveness

Many scholars have been interested in the causes of victory and defeat, broadly defined, in interstate wars. In addition to material factors like military personnel under arms, additional, non-tangible factors have also been hypothesized to affect war outcomes. Some argue that regime type is an important non-material factor, arguing, through varying mechanisms, that democracies are significantly more likely to win wars (Bueno de Mesquita, Morrow, Siverson, and Smith 1999; Choi 2004; Lake 1992; Reiter and Stam 1998). Others argue that key determinants of victory are harmonious civil-military relations (Biddle and Long, 2004; Biddle and Zirkle 1996; Rosen, 1995), human capital (Biddle and Long 2004; Brooks 2003), troop morale(see Reiter and Stam 2002: 60-64), or force employment (Biddle, 2004).

The primary datasets used to test theories of victory and defeat in war rely on categorical measures of outcomes, such as Correlates of War's victory/defeat variable, Stam's (1999) war-level win/lose/draw variable, and Slantchev's (2004) war-level defeat/concessions/gains/victory variable. These measures are appropriate for studies seeking to explain the sources of victory and defeat in the broadest, strategic sense, but are not useful measures of military effectiveness on the battlefield. Effectiveness is a reflection of relative costs and the efficiency with which combatants fight. Highly effective militaries can still fail to achieve their political goals, as demonstrated by the many cases in which asymmetric wars had unexpected final outcomes (e.g. Arreguin-Toft 2001; Mack 1975; Merom 2003; Sullivan 2012).

The primary exception is Grauer and Horowitz (2012) who focus specifically on military effectiveness. However, like the war level datasets, they opt for a dichotomous victory/defeat measure selecting ‘decisive’ battles with post-hoc knowledge of how the conflict ended, thereby deriving important judgments about military effectiveness from that subset of battles. However, a state's military effectiveness cannot be judged by reference to only such decisive battles; effectiveness must include some accounting of costs suffered, and not only at the culminating point of victory, but in the often long, indecisive series of battles that lead to the final outcome. Judging any state at war by the stunning victories and defeats that appear to have turned the tide of the war overlooks a tremendous amount of less interesting, but equally important combat. In addition, focusing on a dichotomous measure of battle ignores the efficiency with which combatants fight and the relative costs of victory. For example, although China won the Lam Son campaign in the Sino-Vietnam War, the battlefield effectiveness of the Vietnamese during the campaign convinced China that it would be too costly to continue onto Hanoi (Grauer and Horowitz 2012: 101-102). It was the ability of the Vietnamese to inflict high costs on the Chinese that mattered, not the outcome of the battle.

These challenges lead us to pursue a more fully informed measure of military effectiveness: Loss Exchange Ratios. LERs provide information on the ability of a military to protect its own soldiers and inflict losses on its enemy, focusing on the process of war fighting, rather than the ultimate outcome.

Focusing on the process has two benefits. First, it provides a continuous variable that captures important differences between combatants that categorical measures of war-level victory and defeat do not. COW codes both the US in the Gulf War and Russia in the Winter War as winners, but their battlefield effectiveness was significantly different and sent very different signals to other states. The US performance was heralded as a demonstration of America's military prowess (Perry 1991) while Russia's costly victory was seen as confirmation that Stalin's purges had crippled the Red Army (Edwards 2008). The LERs from these conflicts capture these differences. The US LER in the Gulf War was less than .03 while Russia's was more than 5 in the Winter War. Similarly, pooling India in the 2nd Kashmir War and Libya in the Anzou War together as war losers ignores the fact that they performed very differently on the battlefield, with India achieving an LER of .86 and Libya suffering an LER of 7.

Having a continuous measure of effectiveness also provides insight into the nebulous 'draw' category used in categorical coding of war outcomes. The results of analyses of war outcomes often pivot around how the draw category is treated. Some treat draws as a distinct category (Downes 2009; Lyall and Wilson 2009; Croco 2011), while others exclude them from the analysis (Reiter and Stam 2002). LERs can shed light on these draws. The War of Attrition may have ended in a draw, but the LERs of the combatants demonstrate that there were meaningful differences in how the adversaries performed, with Israel achieving an LER of 0.07 and Egypt suffering and LER of 13. Having a continuous variable reflecting relative costs thus allows us to explore the process through which war outcomes are achieved.

This nuance also allows us to measure changes in military effectiveness over time. Looking at Austria in the 19th century, we see a steady decline in the five wars Austria fought between the Austro-Sardinian War, in which Austria achieved an LER of 0.5 and the Seven Weeks War, where Austria suffered an LER of 2.85. This matches the general understanding of Austria's decline during that period, but looking only at categorical war outcomes obscures this, as Austria was able to secure a number of high-cost 'victories' after its initial loss in the War of Italian

Unification.

Military effectiveness as a concept separate from war outcomes is critical to questions about war outcomes and the nature of military power. For example, the significance of the Six Day War extended beyond Israel's territorial gains. The efficiency with which Israel fought, and the ease with which it conquered large swaths of Arab land, changed how its adversaries approached Israel and planned for military contingencies. Similarly, it was the extent of the US victory in the Gulf War that prompted discussions of American primacy (Perry 1991). In both cases, the fighting efficiency of the victor was as important as the victory itself. As for questions of military power, knowing how efficiently a state fights is key to evaluating how useful (or threatening) its material resources would be in a contest of arms. Russia's performance in the Winter War convinced Nazi Germany that Russia was incapable of effectively wielding the considerable resources it had its disposal and was thus weak enough to be conquered (Edwards 2008). Prussia's efficiency during the Seven Weeks War convinced France that Prussia was a threat and ultimately led to the Franco-Prussian War (Wawro 2000).

Aggregate war outcomes provide very little information about fighting efficiency because war outcomes are a function of numerous factors. LERs measure efficiency directly by measuring how well a state protects its own soldiers while killing its enemy's. Thus, the LERD offers a replicable, quantitative, continuous measure of military effectiveness, which allows scholars to investigate the sources and consequences of success in warfare by focusing on the efficiency with which combatants fight.

Existing War Casualties Datasets

The major stumbling block preventing scholars from using LERs more widely is the scope of the available data on casualties. Scholars using this concept have relied primarily on data derived from the US Army's CDB-90 dataset, sometimes known as the HERO dataset. As Biddle notes, however, the original CDB-90/HERO data suffer from severe shortcomings that are only partially remedied in a 'cleaned' version of the data used to create LERs by Biddle (2004), Biddle and Long (2004), and Beckley (2010).[1] Because the CDB-90 data have been collected for only a small sample of battles across wars deemed to be historically significant by the original coders, loss-exchange ratios based on these data may not be able to yield generalizable results in studies explaining military effectiveness more broadly. The 381 battles included in the dataset represent only 18 of 46 COW wars in the twentieth century, and 85% of the battles involve the United States, Israel, or Germany (Biddle 2004: 152). The entire Vietnam War is missing with the exception of one battle, for which there is incomplete information. Also problematic is the absence of source lists for the numbers provided by HERO. While a general codebook is available, it is not possible to retrace the researchers' steps to evaluate the reliability or validity of their coding.

Pilster and Boehmelt (2011) attempt to address the missing data problem by expanding the range of wars for which battle-level casualties are available, but they collect information only for the period from 1965 through 1999, and they only use one source for all of their casualty numbers, Clodfelter's (2008) encyclopedic military history.

[1] See Biddle (2004), pp. 152-153 for a description of the problems with the original CDB-90 data.

Since the limited scope of CDB-90/HERO leaves out most interstate wars, many studies using information on casualties rely on war-level data from the Correlates of War project covering interstate wars fought between 1816-2007. [2] However, COW data on casualties are available only in non-directed form, meaning that casualties are tallied for each combatant across the entire war. In wars involving more than two combatant states, this makes it impossible to calculate LERs for each dyadic set of combatants, since we do not know which combatant imposed which casualties on the others. At best, it is possible to calculate allied LERs, but this masks important variation in the efficacy of the different states that make up the alliance. In order to fill the gap in battle-level LER figures, we present and utilize new, directed casualties data for all documented battles in wars involving more than two combatants.

Deriving Loss Exchange Ratios and Constructing the Dataset

Because LERs can be easily calculated for bilateral conflicts, we begin by identifying the multilateral interstate wars from the Correlates of War Interstate War Data (3.0). Following Reiter and Stam (2002), we disaggregate multiphase conflicts into separate wars, treating World War I as five distinct conflicts, World War II as nine different conflicts, and the Vietnam War as two conflicts.[3]

We obtain data on battle-level casualties for these wars using Clodfelter's *Warfare and Armed Conflicts: A Statistical Reference* (2008), Dupuy and Dupuy's *The Harper Encyclopedia of Military History: From 3500 B.C. to the Present*, and over 70

[2] The Peace Research Institute at Oslo (PRIO) has also collected data on casualties. However, their temporal frame only covers 1946-2008 and they do not disaggregate casualties by the parties in the dispute.

[3] See the LERD codebook for a detailed description of how the wars were disaggregated.

university press sources.[4] Limiting our data collection to these major reference works and university press sources helps ensure that the sources are peer reviewed and that their casualty numbers are generally credible.

For each source, we begin by recording every textual reference to battles that provide casualty figures for all participants. For each quote, we record, when available, the name or location of the battle, the battle's end date, the participants, the role of those participants as attacker or defender, casualties for each participant, the LER for each participant, and the type of casualties used to calculate that LER.[5] These data make up the dataset "Source Battle Data," which contains 2,488 directed-dyad battle records. We use these data to create the 'Battle Data,' which averages the combatant's LER and total casualties for each battle across all historical sources. It also records the battle name/location, the end date of the battle, the participants on each side of the battle, and the battle role of those participants as attacker or defender.

The 'Battle Data' set is a useful resource in its own right, but it can also be aggregated into war-level data in a variety of ways. Team data can be generated to compare how different alliance partners performed when working together versus separately; dyadic data can be generated to examine how combatants fared against specific adversaries; types of battles can be selected so separate LERs are computed for the combatant's offensive and defensive battles. The LERD project files include the documentation, all of the datasets, and the aggregation code that we use here. This transparency should make it straightforward for researchers to re-aggregate

[4] Of the 1244 Battle Records, 509 come from Clodfelter and 79 come from Dupuy and Dupuy.

[5] 'Killed in Action' or 'Casualties,' which includes both killed in action and wounded in action.

the data to suit their research design.

Descriptive Statistics

Because the LER is a ratio, it is not normally distributed. In our dataset, roughly half the observations lie between 0 and 1, while the corresponding observations lie between 1 and 645. Because of this, examining the logged transformation of the LER variable is more informative than examining the raw data. The minimum of the logged LER is -2.71, which corresponds to an LER of 0.002. Britain achieved this LER during the destruction of the Vichy French fleet at Mers el Kabir in Algeria, following the French surrender to Nazi Germany. In that engagement, Britain suffered only two fatalities, while inflicting more than 1,297 French fatalities, making it the most lopsided battle in the data set. The maximum LER of 645 (logged value=2.81) is the corresponding LER for France in the battle. The mean of the logged variable is 0, which corresponds to an LER of 1, approximately the LER both Egypt and Israel suffered in Operation Horev during the Palestine War of 1948. The standard deviation of the logged variable is 0.87, thus the 7.24 LER (logged value=0.86) suffered by China at the hands of US forces during Operation Commando during the Korean War, and the 0.13 LER (logged value=-0.86) achieved by the US during Operation Idaho Canyon against the North Vietnamese during the Vietnam War, fall within one standard deviation. Figure 1 displays the distribution of the logged LER variable for the battle-level dataset. The scale has been changed back to the original LER variable to aid in interpretation.

[Insert Figure 1]

Demonstration Analysis

One of the key unanswered questions in the military effectiveness literature is whether democracies are more effective fighters than non-democracies (for example Desch 2008; Downes 2009; Lake 1993; Reiter and Stam 2002). This debate is embedded in a larger debate over whether democracies are more likely to win wars. As such, most of the competing arguments have been tested using outcomes as the key dependent variable in war-level analyses (for example Gelpi and Greisdorft 2001; Reiter and Stam 2002). Because war outcomes are not synonymous with military effectiveness, testing these theories using the LER data provide a new resource by which to evaluate the mechanisms underpinning the theories of democratic victory and directly test the theories relating to military effectiveness. Below, we compare Reiter and Stam's (2002, 2009) and Downes' (2009) models using both the categorical indicators used in their initial studies and the LERD's war level LER variable to demonstrate how the LERD can contribute to these important debates.

Lake found that between 1816 and 1988, democracies won 80% of the wars they participated in, while autocracies only won 43% (Lake 1992: 31). Since then, scholars have offered a number of explanations for this pattern, including democracies' tendency to select themselves into easier wars (Bueno de Mesquita et al 1999), their ability to commit more resources to the battlefield (Lake 1992), and their ability to attract other democratic allies (Choi 2004). In a series of articles that was later compiled into a book, Reiter and Stam (2002) investigated each of these mechanisms. They concluded that democracies tend to win the wars they fight because they select themselves into easier conflicts and because they are more proficient on the battlefield. Their findings have subsequently been questioned by a number of scholars, most recently Downes (2009) who critiques not only the theoretical foundations of their findings, but also questions the overall correlation between democracy and victory.

We contribute to the debate by applying Reiter and Stam's models to military effectiveness, allowing us to evaluate the two mechanisms that they propose to explain the connection between democracy and victory in war: opponent selection and military proficiency on the battlefield. If these mechanisms are correct, democracies should achieve more favorable LERs because their proficiency on the battlefield enables them to simultaneously kill their enemies while protecting their own soldiers, an effect exacerbated by the fact that they choose weak opponents who should be easy to neutralize on the battlefield.

We test these theories using a war-level measure of LER as our dependent variable. The LER is defined as the combatant's casualties divided by enemy casualties. Countries that limit their own casualties while inflicting high losses on the enemy will have low LERs. Thus, a low LER is associated with military effectiveness. We use the log of the combatant's LER, which is calculated using the Correlates of War (3.0) casualty data for bilateral conflicts and aggregated LER's from the LERD for multilateral conflicts. We compute LERs for combatants involved in multilateral by taking the weighted average of the combatant's LER for every battle the combatant was involved in, using the battles' total casualties to assign weights.[6]

An initial examination of the LER data suggests that democracies are not more effective fighters. Figure 2 compares the distribution of LERs across democracies, anocracies, and autocracies.[7] Though the mean LER for democratic combatants is lower than for anocratic and autocratic combatants, they all center closely around a logged LER of 0 (LER=1). Furthermore, the variance is quite high for all combatants and even more so for democracies. The vast majority of observations for both anocracies and autocracies fall within one standard deviation of the mean score for democratic combatants and the tail end of the confidence interval for democratic combatants lies above (and so is indicative of worse battle field performance) the upper end of the threshold for non-democracies.

[Insert Figure 2 Here)

Figure 3 explores this relationship in more depth, showing how LER varies across the entire range of polity scores for initiators, targets, and joiners. Contrary to the theories espoused by Reiter and Stam, there does not appear to be a consistent downward trend in LER as a countries polity score rises. The black line shows the lowess curve for all war participants: it is nearly flat. When participants are disaggregated by role, we see some variation in the patterns but the relationship between democracy and effectiveness is weak, at best. Although there is a slight

[6] For example, when computing Israel's LER for the Six Day War, Israel's LER for the battle of Rafah is weighted much more heavily than its LER for the battle of Ammunition Hill because the total casualties for Rafah were much higher (2,500 Israeli and Egyptian KIA) than Ammunition Hill (106 Israeli and Jordanian KIA).

[7] Democracies are countries with a polity score above 17. Autocracies are countries with a polity score below 5. Anocracies have polity scores between 5 and 17.

downward trend for initiators, the lowess curve for targets is flat, and there is no discernable pattern for joiners.

[Insert Figure 3 Here]

Its possible that this aggregated data mask important trends in the data and that by controlling for other predictors of military effectiveness, the relationship between democracy and a combatant's LER could be strengthened in a way that confirms the hypotheses of theories of democratic effectiveness. We investigate this using the most comprehensive models provided by Reiter and Stam (2002, 2009) and their most recent challenger, Downes (2009).

Reiter and Stam (2002) use a probit model with robust standard errors to estimate the effect of regime type on a dichotomous win/lose variable. Draws are excluded from the analysis. Their key independent variables are *politics*initiation* and *politics*target*, which are the POLITY Scores for initiators and targets, respectively.[8] Table 2.2, Model 4 is their most comprehensive model, including control variables for material capabilities, troop quality, strategy, terrain features, and the interaction between strategy and terrain (Reiter and Stam 2002: 45).[9] Although this model was developed to predict war outcomes, these variables should all influence battlefield effectiveness and thus provide a more stringent test of the democratic effectiveness hypothesis when LER is used to measure effectiveness. Countries armed with better weapons typically have greater firepower at their disposal and more protective gear: allowing them to protect their own soldiers

[8] They do not include the constitutive *Politics* term because they only have two categories of states and including it would lead to multicollinearity problems (Reiter and Stam 2002: 40-41).

[9] Details on these variables can be found in Reiter and Stam (2002: 38-44).

while killing more of the enemy. Biddle (2004) found that one of the key determinants of effectiveness was the ability of combatants to employ the modern system. Thus, both strategy and troop quality should be decisive predictors; the former because of its relation to force employment and the latter because implementation of the modern system requires highly trained troops. Finally, because terrain has a direct effect on both the effectiveness of differing strategies and can also have a direct effect on the combatants' ability to protect their soldiers while targeting the enemy, both the constitutive terrain term and the interactive effects could be important predictors LER.[10]

[Insert Table 1 Here]

Table 1, Model 1 provides the results of Reiter and Stam's model. All statistical results were created using Stata 11 software. The two key variables, *politics*initiation* and *politics*target,* are positive and statistically significant, suggesting that both democratic initiators and democratic targets are more likely to win the wars they fight.

Model 3 replicates the analysis using standard regression with the combatants' LER as the dependent variable. Following Reiter and Stam, we used robust standard errors to account for heteoskedasticity.[11] As discussed above, low LERs are associated with military effectiveness. If democracies are more effective

[10] There are other important predictors of effectiveness that are not included in this model and there are alternative ways of measuring terrain and strategy. However, the goal of this paper is to replicate Reiter and Stam's analysis using new data. Given this goal, we focus on the variables they include to test their hypotheses.

[11] In order to maintain consistency with Reiter and Stam's models, the reported errors are not clustered on war. Because observations within wars are not independent, we ran robustness tests using clustered errors. This did not change our substantive results.

fighters we should see LERs decline as the polity score (in this case the *politics* interaction terms) increase.

The *politics* interactive terms are both insignificant in Model 3, suggesting that neither democratic initiators nor democratic targets are more effective at fighting. Looking at Model 2, which estimates Reiter and Stam's probit model on the same cases for which the LER is available, we can see that the differing results are partially due to a change in the sample of cases being evaluated. The *politics*initiation* variable is insignificant in Model 2 and Model 3. However, the *politics*target* variable remains significant in Model 2. It only becomes insignificant when the new dependent variable is used in Model 3, suggesting that the results are driven by the nature of the dependent variable, not sample selection.[12] The null result for the politics variables are all the more striking because other key variables continue to behave as expected: combatants with more capabilities at their disposal with higher quality troops are more effective, but those operating in rough terrain are less so.

Downes presents an alternative model for predicting war outcomes: one which yields very different results. Unlike Reiter and Stam, he includes draws as a potential outcome and consequently uses an ordered probit model to estimate the effect of democracy on war outcome, using clustered robust standard errors to

[12] A Shapiro Wilks test shows that we can reject the null hypothesis that the residuals are normally distributed (p=.054), but visual tests of the residuals demonstrate that the data closely mirror a normal distribution. We use White's test for generalized heteoskedasticty and find that the error terms are not homogenous (p=.001). This confirms our decision to use robust standard errors (and in later models to cluster those errors on war). We also ran robustness tests to evaluate how sensitive the results were to outliers, identifying three likely outliers (Germany and Yugoslavia in the Yugoslav theatre of WWII and Jordan during the Palestinian War) and three potential outliers (Greece during WWI, Israel during the War of Attrition, and Russia during the Sino-Soviet War). Exclusion of these cases individually or together did not affect the results.

account for the interdependence of observations from within the same war. His analysis treats joiners as a separate category, rather than coding them as initiators or targets, and he specifies the interaction between the politics variable and war role variables in a more traditional manner, including the constitutive *polity*, *initiation* and *target* variables, as well as their interactions.[13] With this specification, the effect of democracy is measured by the polity*initiation variable for initiators, the polity*target variable for targets, and the constitutive polity term for joiners. He includes all of Reiter and Stam's control variables. Downes' results are reported in Model 4, below.

[Insert Table 2 Here]

None of the *polity* variables are significant, suggesting that neither democratic initiators, targets, nor joiners are more likely to win the wars they fight. This finding is confirmed for military effectiveness, in Model 6, where the analysis is re-estimated using regression analysis with LER as the dependent variable.[14]

In replying to Downes' critique, Reiter and Stam (2009) argue that focusing on a linear relationship is problematic because mixed regimes are especially likely to lose wars compared to both democracies and autocracies. It is plausible that regime type might also have a non-linear affect on effectiveness: for example, if anocratic regimes have particularly toxic civil-military relations or if individualist norms have not been adequately internalized. To test this hypothesis, we replicate Model 1 in

[13] He also rescales the *polity* variable to 1-21 rather than -10 to 10.

[14] As with Reiter and Stam's model, the assumption of normally distributed errors does not hold. The Shapiro Wilks test shows we can reject the null hypothesis (p=.046). Visual tests confirm, however, that the distribution closely approximates a normal distribution. This model assumes heteroskadistic errors clustered on war. An analysis of the homoskedasticity assumption using an identical model with non-clustered errors supports this decision (p=.018 for White's test). The results from this model are insensitive to the exclusion of outliers.

Reiter and Stam's (2009) reply to Downes, 'Another Skirmish over Democracies.' This model tests for a curvilinear relationship between war outcomes and democracy for initiators using fractional polynomials. The model is identical to Model 1 reported above, except that they replace the *politics*initiation* variable with two fractional polynomials: *poly pol 1* is defined as $x^{-1/2}$ and *poly pol 2* is defined as $x^{-1/2}(\ln(x))$ where x=(politics*initiation+11)/10. Both terms should be significant if extremely low levels of democracy are moderately correlated with positive war outcomes, mixed regimes are negatively correlated with positive war outcomes, and high levels of democracy are highly correlated with positive war outcomes or the LERs of mixed regimes are considerably higher than both democracies and autocracies.[15]

[Insert Table 3 Here]

Reiter and Stam find a curvilinear relationship between regime type and war outcome for war initiators, but again, when this model is re-estimated using LER as the dependent variable in regression analysis in Model 9, this effect disappears. None of the regime type variables have a statistically significant effect on military effectiveness as measured by the combatants' LER and an F-test confirms that the joint effect of the polynomial terms cannot be distinguished from 0 (p=.736).[16]

Thus, when military effectiveness is examined, rather than war outcomes, democracy does not have a statistically significant effect. None of the regime type

[15] See Reiter and Stam (2002: 41).

[16] The Shapiro Wilks test shows that we cannot reject the null hypothesis of normally distributed errors (p=.331) though there is evidence of heteroskedasticity (p value for White's test is 0.0036). That is to be expected, given the interdepence between observations from the same war. The use of robust clustered errors alleviates this problem. As before, these results are insensitive to the exclusion of outliers.

variables are statistically significant and the model fit is fairly low, accounting for a little more than 30% of the variation in combatants' LER. While this does not settle the debate on regime type and war outcomes, it does demonstrate that democracies are not particularly effective on the battlefield. When effectiveness is measured by considering how efficiently combatants fight, democracies do not have an edge on their autocratic, or anocratic, counterparts. This null result is remarkably robust against a number of modeling choices: the exclusion of draws, different treatment of joiners, linear and curvilinear models, the exclusion of outliers, and the inclusion of clustered errors.

This analysis demonstrates how the LERD can be used to further our understanding of the sources of military effectiveness. Although the models Reiter and Stam and Downes employ were designed to test the effect of democracy on war outcomes, the theories they develop have observable implications for the relationship we should see between democracy and battlefield military effectiveness, which is what the LER measures. If Reiter and Stam are correct that democracies select themselves into easier wars and then fight more proficiently on the battlefield, democracies should have lower LERs. The war-level LERs are essentially aggregated battle level statistics and speak directly to the question of fighting proficiency and if democracies pick inept opponents, it should be easier for them to protect their own soldiers while killing the enemy. If Downes is correct that domestic political calculations do not necessarily lead democratic leaders to choose 'easy' wars and at times actually create incentives for leaders to initiate conflict when the risks are high then there should be no correlation between low LERs and democracy. The distribution of democratic adversaries should be random and democracies should not find it particularly easy to kill their enemies or protect their own soldiers just because they selected weak opponents. Testing these implications using new data moves the debate forward, away from technical modeling choices (such as whether to include draws) toward an evaluation of the key theoretical mechanism. It also creates incentives for scholars to develop theories of effectiveness that can explain variation in both war outcomes and fighting efficiency.

We hope that the LERD can contribute to other debates about military effectiveness in a similar manner. One of the benefits of the LERD is that it has the potential to do so at multiple levels of analysis. The above discussion focuses on war-level tests of the democratic victory hypothesis, but these theories have also been tested at the battle level (Biddle and Long 2004; Grauer and Horowitz 2012; Reiter and Stam 1998). The LERD provides a great resource with which scholars can extend these analyses to a wider range battles.

Conclusion

While the replications we perform here provide an interesting demonstration of the utility of the LERD, they are only the beginning of what could be done. Scholars can use the LERD to revisit some of the classic questions of the military effectiveness literature and expand their questions into new areas. The LERD makes it possible to evaluate directed hypotheses about the effects of political and strategic factors at the war and battle level. Since many of the major questions of the military effectiveness literature are by nature directed, this should make it possible to make significant improvements in our knowledge of factors improving or worsening military effectiveness.

We also hope that these data may serve as a cross-national, cross-sectional measure of demonstrated, as compared to potential, military power. Much of the international relations literature treats material capabilities as synonymous with military power, and in the quantitative international relations literature, power is almost always measured with respect to a state's material resources despite the fact that these measures are not great predictors of actual military outcomes (Tellis, Bially, Layne, and McPherson 2000: 30-31).

If power is the ability to influence events, military power is *the ability to influence battlefield outcomes*. We believe that a state's historical LERs may measure this more accurately than variables measuring material capabilities. Historical LERs capture the demonstrated ability of a state to kill the enemy and protect its own soldiers. They incorporate both the material resources a state brings to the battlefield and non-tangibles such as force employment, strategic leadership, and adaptability to measure a state's ability achieve favorable battlefield results. Using states' past LERs to measure military power could enable researchers to test whether the demonstrated ability to perform on the battlefield affects the degree to which states can influence the behavior of others in the international system, as predicted by Waltz (1979) and others. Perhaps states that perform well are better able to deter potential opponents from taking threatening actions. Perhaps states that perform well attract allies while those that perform poorly are left without friends when the war horn calls. Perhaps states with effective militaries are able to secure better deals at the negotiating table. Using the LER to measure military power would enable scholars to test these theories by providing a means by which states can be identified and ranked with reference to their demonstrated ability to fight effectively. Thus, the LERD has the potential to provide a useful new resource, not only for scholars studying the sources of military effectiveness, but also those studying military power more broadly.

References

Arreguin-Toft, Ivan. (2001) How the Weak Win Wars: A Theory of Asymmetric Conflict. *International Security* 26(1):93-128.

Beckley, Michael. (2010) Economic Development and Military Effectiveness. *Journal of Strategic Studies* 33(1):43-79.

Biddle, Stephen. (2004) *Military Power: Explaining Victory and Defeat in Modern Battle.* Princeton: Princeton University Press.

Biddle, Stephen and Robert Zirkle. (1996) Technology, Civil-Military Relations, and Warfare in the Developing World. *Journal of Strategic Studies* 19(2):171-212.

Biddle, Stephen and Stephen Long. (2004) Democracy and Military Effectiveness – A Deeper Look. *Journal of Conflict Resolution* 48(4):525-546.

Brooks, Risa A. (2003) Making Military Might: Why Do States Fail and Succeed? A Review Essay. *International Security* 28(2):149-191.

Bueno de Mesquita, Bruce, James D Morrow, Randolph M Siverson, and Alastair Smith. (1999) An Institutional Explanation of the Democratic Peace. *American Political Science Review* 93(4):791-807.

Ajin Choi. (2004) Democratic Synergy and Victory in War, 1816-1992. *International Studies Quarterly* 48(3):663-682.

Clodfelter, Michael. (2008) *Warfare and Armed Conflicts: A Statistical Encyclopedia of Casualty and Other Figures, 1494-2007.* Jefferson, NC: McFarland & Company.

Correlates of War Project. National Material Capabilities (v3.02). Accessed December 13, 2013 at http://www.correlatesofwar.org/COW2%20Data/Capabilities/nmc3-02.htm.

Croco, Sarah E. (2011) The Decider's Dilemma: Leader Culpability, War Outcomes, and Domestic Punishment. *American Political Science Review* 105(03):457-477.

Desch, Michael C. (2008) *Power and Military Effectiveness: The Fallacy of Democratic Triumphalism.* Baltimore: Johns Hopkins University Press.

Downes, Alexander B. (2009) How Smart and Tough are Democracies? Reassessing Theories of Democratic Victory in War. *International Security* 33(4):9-51.

Edwards, Robert. (2008) *The Winter War: Russia's Invasion of Finland, 1939-1940.* New York: Pegasus Books.

Gelpi, Christopher F. and Michael Griesdorf. (2001) Winners or Losers? Democracies in International Crisis. *American Political Science Review* 95(3):633-647.

Grauer, Ryan and Michael C. Horowitz. (2012) What Determines Military Victory? Testing the Modern System. *Security Studies* 21(1):83-112.

Lake, David A. (1992) Powerful Pacifists: Democratic States and War. *American Political Science Review* 86(1):24-37.

Lyall, Jason and Isaiah Wilson III. (2009) Rage Against the Machines: Explaining Outcomes in Counterinsurgency Wars. *International Organization*, 63(1):67-106.

Mack, Andrew. (1975) Why Big Nations Lose Small Wars: The Politics of Asymmetric Conflict. *World Politics* 27(2):175-200.

Merom, Gil. (2003) *How Democracies Lose Small Wars: State, Society, and the Failures of France in Algeria, Israel in Lebanon, and the United States in Vietnam.* New York: Cambridge University Press.

Perry, William J. (1991) Desert Storm and Deterrence. *Foreign Affairs* 70 (4).

Pilster, Ulrich and Tobias Boehmelt. (2011) Do Democracies Engage in Less Coup-Proofing? On the Relationship Between Regime Type and Civil-Military Relations. *Foreign Policy Analysis* 8(4):355-372.

Reiter, Dan and Alan C Stam III. (1998) Democracy and Battlefield Military Effectiveness. *Journal of Conflict Resolution* 42(3):259-277.

Reiter, Dan and Allan C Stam. (2002) *Democracies at War*. Princeton: Princeton University Press.

Reiter Dan and Allan C Stam. (2009) Correspondence: Another Skirmish in the Battle over Democracies and War. *International Security* 34(2):194-200.

Rosen, Stephen Peter. (1995) Military Effectiveness: Why Society Matters. *International Security* 19(4):5-31.

Slantchev, Branislav L. (2004) How Initiators End Their Wars: The Duration of Warfare and the Terms of Peace*.* *American Journal of Political Science* 48(4):813-829.

Stam, Allan C III. (1999) *Win, Lose, or Draw: Domestic Politics and the Crucible of War.* Ann Arbor: University of Michigan Press.

Sullivan, Patricia. (2012) *Who Wins? Predicting Strategic Success and Failure in Armed Conflict.* New York: Oxford University Press.

Tellis, Ashley J, Janice Bially, Christopher Layne, and Melissa McPherson. (2000) *Measuring National Power in the Postindustrial Age.* Santa Monica: Rand.

Waltz, Kenneth N. (1979) *Theory of International Politics.* Boston: McGraw Hill.

Wawro, Geoffrey. (2000) *Warfare and Society in Europe 1792-1914*. New York: Routledge.

Tables and figures

Figure 1. Histogram of Battle Level LERs

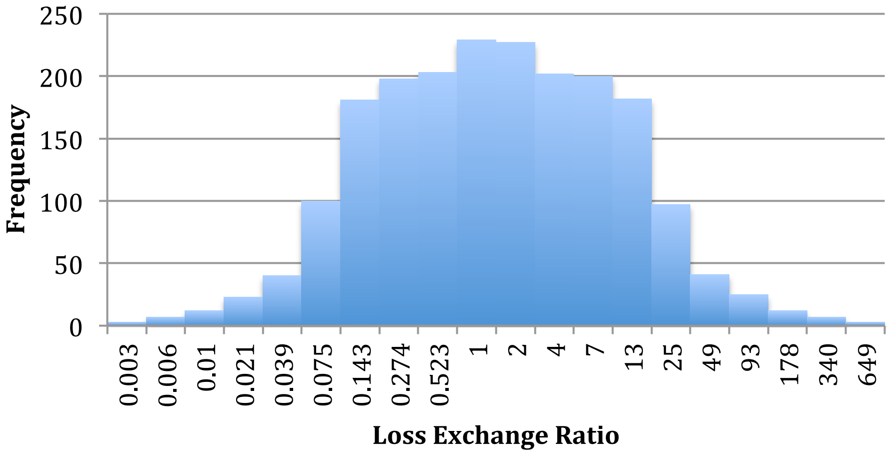

Figure 2. Boxplot of LER by Regime Type

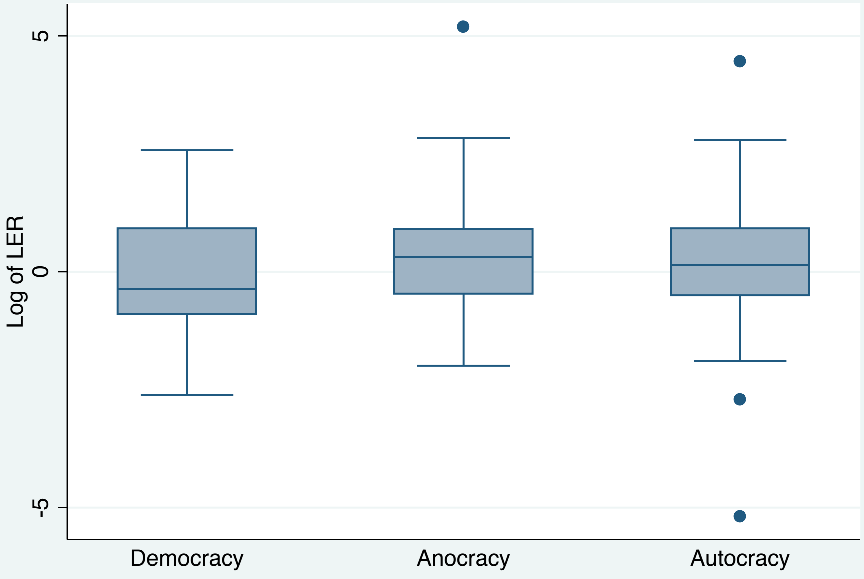

Figure 3.Scatter Plot of Loss Exchange Ratio by Polity Score

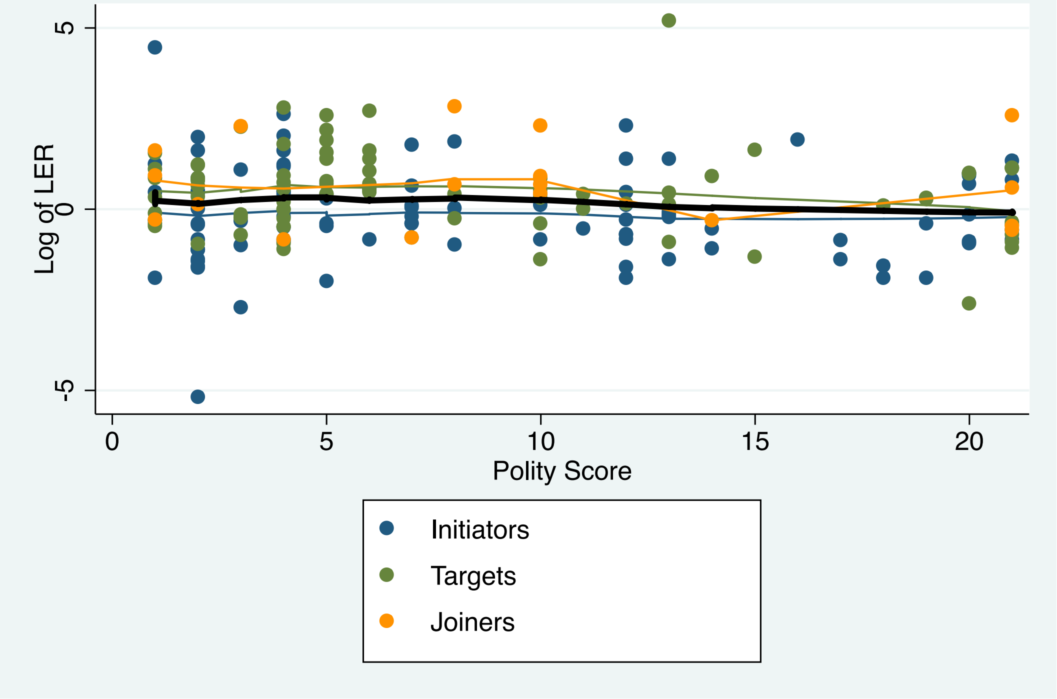

Table 1. Replication of Table 2.2, Model 4 from *Democracies at War*			
	Model 1 Replication (Probit)	Model 2 Replication with Cases for LER Analysis (Probit)	Model 3 Regression with LER
*Politics*Initiation*	0.067** *0.03*	0.04 *0.03*	-0.01 *0.02*
*Politics*Target*	0.064** *0.03*	0.078** *0.03*	-0.02 *0.02*
Initiation	0.914*** *0.34*	1.074*** *0.40*	-0.406* *0.23*
Capabilities	3.727*** *0.53*	3.852*** *0.55*	-0.708** *0.32*
Alliance Capabilities	4.722*** *0.68*	5.418*** *0.80*	-0.862* *0.47*
Troop Quality	0.05 *0.03*	0.07 *0.04*	-0.040*** *0.01*
Terrain	-10.932*** *2.94*	-12.365*** *3.38*	5.062*** *1.87*
*Strategy*Terrain*	3.560*** *0.97*	4.098*** *1.13*	-1.473** *0.59*
Strategy 1	7.235**	9.006***	-2.50

	2.89	*3.37*	*1.74*
Strategy 2	3.479*	4.373*	-1.47
	1.99	*2.25*	*1.26*
Strategy 3	3.357**	4.146**	-1.10
	1.43	*1.61*	*0.87*
Strategy 4	3.069**	3.779**	0.07
	1.25	*1.47*	*0.64*
Constant	-5.517***	-6.701***	1.43
	1.70	*1.97*	*0.97*
N	197	161	161
[Psuedo] R2	0.52	0.55	0.33
*Robust standard errors given in italics: * p<.1, ** p<.05, ** p<.01*			

Table 2. Replication of Model 3 in "How Smart and Tough are Democracies?"

	Model 4 Replication (Ordered Probit)	Model 5 Replication with Cases for LER Analysis (Ordered Probit)	Model 6 Regression with LER
Polity 21	0.03 *0.05*	0.02 *0.05*	0.01 *0.04*
Initiation	0.37 *0.56*	0.36 *0.64*	-0.63 *0.61*
Target	-0.16 *0.60*	-0.27 *0.59*	-0.14 *0.58*
*Polity*Initiation*	-0.01 *0.05*	0.00 *0.05*	-0.02 *0.05*
*Polity*Target*	-0.01 *0.05*	0.00 *0.05*	-0.04 *0.04*
Capabilities	2.354*** *0.49*	2.267*** *0.50*	-0.662* *0.36*
Alliance Capabilities	3.000*** *0.74*	3.059*** *0.74*	-0.804* *0.47*
Troop Quality	0.042* *0.02*	0.05 *0.03*	-0.0426*** *0.01*
Terrain	-1.891* *1.13*	-1.71 *1.12*	2.370* *1.33*
*Strategy*Terrain*	0.50 *0.36*	0.49 *0.36*	-0.66 *0.44*
Strategy 1	-0.47 *1.30*	-0.49 *1.32*	-0.21 *1.33*
Strategy 2	-2.554** *0.88*	-2.604*** *0.92*	0.49 *0.96*
Strategy 3	-0.21 *0.65*	-0.21 *0.66*	-0.09 *0.66*
Strategy 4	1.11 *0.73*	0.98 *0.80*	0.67 *0.56*
Constant1	0.80 *0.87*	0.78 *0.90*	0.87 *0.94*
Constant 2	1.44 *0.84*	1.46 *0.86*	
N	233	192	192
[Psuedo R2]	0.2914	0.2865	0.3102

*Robust standard errors given in italics: * p<.1, ** p<.05, ** p<.01*

Table 3: Replication of Model 1 in "Another Skirmish Over Democracies"

	Model 7 Replication (Probit)	Model 8 Replication with Cases for LER Analysis (Probit)	Model 9 Regression with LER
*Politics * Target*	0.01 *0.02*	0.02 *0.02*	-0.02 *0.01*
Initiation	6.0982*** *1.76*	5.108** *2.14*	0.14 *1.73*
Poly Pol 1 (first curvileanar term)	-4.735*** *1.66*	-3.675* *2.07*	-0.20 *1.23*
Poly Pol 2 (second curvilinear term)	-4.711*** *1.68*	-3.606* *1.97*	-0.98 *1.77*
Capabilities	3.902*** *0.74*	3.949*** *0.76*	-0.830** *0.39*
Alliance Capabilities	4.787*** *1.10*	5.360*** *1.23*	-0.923* *0.54*
Troop Quality	0.05 *0.04*	0.07 *0.05*	-0.039*** *0.01*
Terrain	-13.698*** *3.71*	-14.360*** *4.09*	4.486** *2.06*
*Strategy*Terrain*	4.370*** *1.21*	4.604*** *1.34*	-1.275* *0.65*
Strategy 1	9.521*** *3.29*	10.315*** *3.76*	-2.11 *1.79*
Strategy 2	4.434* *2.50*	4.686* *2.70*	-0.89 *1.42*
Strategy 3	4.511*** *1.61*	4.798*** *1.81*	-0.94 *0.89*
Strategy 4	3.676*** *1.28*	3.907*** *1.49*	0.08 *0.63*
Constant	-6.83 *1.83*	-7.370*** *2.08*	1.49 *0.96*
N	196	160	160
Psuedo R2	0.53	0.55	0.33

*Robust standard errors given in italics: * p<.1, ** p<.05, ** p<.01*