

Core Problems in International Data Collection*

Kristian Skrede Gleditsch

Department of Government, University of Essex &

Centre for the Study of Civil War, PRIO

Wivenhoe Park

Colchester, Essex CO4 3SQ, UK

Tel: +44 (0)1206 872517, Fax: +44 (0)1206 873234

Email: ksg@essex.ac.uk

Kyle Clark Beardsley

Department of Political Science, Emory University

327 Tarbuton Hall

1555 Dickey Drive

Atlanta, GA 30322, USA

Tel: + 1 (404) 727-7506, Fax: +1 (404) 727-4586

Email: kyle.beardsley@emory.edu

Word count: 9,544

* Prepared for *The International Studies Compendium Project, Section for Scientific Study of International Processes*, edited by Paul F. Diehl and James Morrow (general project editor Robert A. Denemark). We are grateful to Paul F. Diehl and James Morrow for helpful comments and discussions. Gleditsch was supported by grants from the Research Council of Norway (NFR-180441) and the UK Economic and Social Research Council (RES-062-23-0259).

Introduction: Data and Inference in Descriptive Questions about World Politics

Many of the central questions of interest to researchers and a general audience about World Politics are inherently descriptive. Some prominent examples include: Has there been a decline in conflict in the international system? Was the 19th century more peaceful than the 20th century? How common are democratic institutions around the world, and how has the extent of democracy changed over time? Does there appear to be any relationship between such changes or “waves of democracy” and conflict in the international system?

Descriptive questions like these should in principle be answerable from empirical data or information about the relevant features. The turn to greater use of scientific methods in the study of international relations, exemplified by the Scientific Study of International Processes section of the International Studies Association, is premised on how attention to the canons of scientific inquiry, systematic data collection, and empirical testing of propositions can help improve our understanding of World Politics. Following the behavioral revolution in the social sciences in the 1960s, we have indeed seen the development of several databases that attempt to take stock of core features relevant to international relations such as conflict between states and democratic institutions. Given the development of such data sources, one might expect that answering the questions posed above should be relatively straightforward, and that we could reach clear and uncontroversial answers to these questions simply by looking at the data at hand.

The development of data on characteristics relevant to the study of international relations has undeniably allowed a great deal of progress to be made on many research questions. However, trying to answer seemingly simple descriptive questions such as those posed above will often alert us to how data rarely speak entirely for themselves. The specific ways in which we pose questions or try to reach answers will often influence our conclusions. Likewise, the specific manner in which the data have been collected will often have

implications for our inferences. answer to core descriptive questions such as those outlined above will often end up being contested by other researchers. Many empirical debates in the study of international relations, upon closer inspection, often hinge on assumptions and criteria that are not made fully explicit in studies based on empirical data. In this chapter, we try to clarify some of the central issues and survey how key problems in international data collection and their interpretation can affect our efforts to answer key descriptive questions about war and peace in the world and its relationship to other features of interest.

A Case Study: Conflict after the Cold War

To use a specific example for illustration, consider the issue of how common conflict is in the world and whether there are any discernable trends in the frequency of conflict over time.

Many prominent scholars had originally predicted that the end of the Cold War would unleash a period of instability with a heightened risk of new conflicts (see, e.g., Mearsheimer 1990 for a particularly pessimistic view; Mueller 1994 provides an interesting review of such predictions). However, an early empirical analysis of the so-called Uppsala Armed Conflict Data (ACD) project collecting information on conflict incidence for the years 1989–94 actually indicated that there appeared to be a decline in new conflicts and fewer ongoing conflicts in the world after the Cold War (see Wallensteen and Sollenberg 1995). This finding of a so-called post-Cold War dip in the frequency of armed conflict has later been replicated by other researchers, including researchers using different data sources on conflict incidence (Gurr 2000), and has gradually become accepted among many conflict researchers (see, e.g., Mueller 2004). The post-Cold War dip in conflict also provides a rare example of an

empirical finding that has managed to get disseminated beyond the confines of academia and conflict studies (see Goldstein 2002; Mack 2002).¹

This decline in the number of conflicts noted in the Uppsala data is undoubtedly an empirical finding, which can be derived directly from these data by simply counting the number of conflicts assigned proper names. However, other researchers have disputed the existence of such a post-Cold War dip in conflict, seemingly also basing their conclusions on empirical evidence and data.² Sarkees, Wayman, and Singer (2003), for example, focusing on the number of casualties in conflicts (from another data source) rather than the number of conflicts, argue that data on conflict “reflect a disquieting constancy in warfare” (p. 49) over the past one hundred and fifty years. Hewitt, Wilkenfeld, and Gurr (2007) noted that the number of countries participating in armed conflict had not fallen to the same extent, and questioned the significance in the observed decline in the number of conflicts if present conflicts were particularly extensive in scope measured by number of participants.

¹ An early op-ed piece in the *Los Angeles Times* on this phenomenon (Wilson III and Gurr 1999) is rumored to have caught the attention of United Nations Secretary General Kofi Anan and in turn spurred the initiative for the subsequent *Human Security Report* (2005). The publication of the *HSR* and its discussion of the post Cold War dip in conflict and its possible causes gave rise to a large number of op-ed pieces in newspapers worldwide.

² The idea of a post-Cold War decline in conflict has also met skepticism from commentators in the popular media (see, e.g., Kaplan 2006). Many argue that “new wars” and forms of violence such as terrorism and genocide have replaced traditional forms interstate conflict (Kaldor 2006). Although such claims are rarely supported using systematic data, this is a potentially valid objection since most existing conflict data require that a state must be one of the actors and require all parties to be a cohesive group to be counted (see, e.g., Gleditsch et al. 2002; Sambanis 2004; Small and Singer 1982). Hence, they would not include other forms of violence and may be inappropriate to consider such displacement effects. However, systematic studies using other data lend no support to claims that genocide and terrorism have become more widespread or severe over time (e.g., Clauset, Young, and Gleditsch 2007; Kalyvas 2001; Rummel 1995)

The fact that researchers draw such different conclusions about what would seem a simple descriptive question about the extent of conflict is bound to seem puzzling to many observers. Upon closer inspection it is possible to find that the conclusions in these studies depend on some obvious differences in research design as well as some more subtle differences in assumptions about how to analyze the available data and the specific manner in which these data have been collected. In this particular case, the evidence for the dip in violence in Wallensteen and Sollenberg (1995) is based on counting the number of unique conflicts in the ACD data. By contrast, the Sarkees et al. argument about the alarming constancy of war is based on the absence of a clear linear trend in a regression of fatalities from conflict on time, whereas Hewitt et al.'s measure does not decline as fast as the number of conflicts since many international interventions in recent conflicts such as Kosovo and Afghanistan have involved large international coalitions. These are very different types of data and measures of the frequency/severity of conflict, and there is no inherent reason why studies using different criteria should need to yield similar conclusions about how common conflict is in the world.

This short case study should impress upon us how many answers to seemingly descriptive question almost inevitably require a number of important additional assumptions, and how it rarely will be possible to simply directly look at available data and reach answers that will be universally accepted as valid. Our argument here is not that it is inherently impossible to answer such questions, or that all approaches are arbitrary and equally valid ways of answering a question. We may certainly legitimately question whether it is reasonable to use the number of participants by very inclusive criteria, including UN operations, as a measure of the scope of conflict³ and whether fitting a linear time trend is a

³ For example, based on the number of participants in the “coalition of the willing” (36), the Iraq War is “larger” than World War I (32) (see Gleditsch 2008). By most other scale criteria (but not duration), it is much smaller.

good way to examine variation in the severity of conflict over time.⁴ However, there may not be any natural or inherently valid empirical measures, and debates on such questions must clarify the potentially contentious issues in collecting and analyzing data. Researchers must be as explicit as possible about the assumptions entailed in their answers and be prepared to defend these before we can meaningfully evaluate differences in their conclusions and their validity.

Core Issues in Data Collection

We will by necessity have to be somewhat selective in the specific issues that we can cover in this Chapter, and we will focus on problems of collecting data on core features that lie at the heart of the discipline, such as how to identify states and conflict between them. However, many of the issues that we discuss are rather general and likely to be relevant for a wider range of data collection projects and research questions in International Relations.

Before we turn to the specifics of particular data collection efforts, it will be helpful to introduce some general criteria and objectives that we would wish data collections to achieve or approach as far as possible. In general, we would prefer data to be collected in a manner that maximizes *content validity* and *objectivity*. Empirical measures have content validity to the extent that the resulting observable indicators reflect the theoretical concepts that researchers are interested in (see Campbell and Stanley 1963). One way to think of objectivity here is as the degree to which the explicit operational definition can be applied in an *intersubjective* manner.⁵ Data become more subjective if two researchers may disagree on

⁴ Lacina, Gleditsch, and Russett (2006) challenge the conclusions of Sarkees et al. (2003), as well as their data. They argue that a curvilinear specification suggests a clear declining level of fatalities of conflict after 1945.

⁵ See, e.g., Vanhanen (1990) and Bollen (1990) for a discussion of objective and subjective measures and indicators in the study of democracy.

how individual observations should be classified or how the criteria should be applied to actual cases. Moreover, we would want our data to be as valid as possible for *allowing comparisons* between individual observations or countries and over time (e.g., King et al. 2003; Summers and Heston 1988). Stated differently, we should be concerned if our data produce relative measures or rankings of observations that fail obvious construct validity tests for a given time slice, or use operational criteria that change over time in ways that may influence our conclusions. The goals of objectivity and content validity can sometimes present very real tradeoffs in data collection efforts. Coding schemes that might be difficult to apply in an intersubjective manner, such as the definition of a state based on relative autonomy or the definition of a crisis based on perception of threat, are often designed to increase the content validity and maximize the ability to draw meaningful comparisons across units. That is, apparent ease of coding consistency is not of much help in itself unless the measures correspond well to the underlying theoretical constructs. When choosing coding rules, one must try to maximize both objectivity and content validity, even though it may be difficult to fully preserve both at the same time.

Finally, just as important as potential problem in data collection are the problems arising from *unobservable or non-random missing data* in comparisons (Rubin 1976) and what we call “*denominator effects*” (Dixon and Boswell 1996; Firebaugh 1992). In particular, we should be very skeptical of descriptive claims if there are likely to be major problems with observing the phenomenon of interest in particular circumstances, or if descriptive claims are based on shares or proportions where the denominator may leave out important relevant actors or individuals. We will return to these issues in the specific applications that we discuss.

The units: How many states or actors?

Identifying the set of the relevant actors or units in the field could be seen as the most fundamental issue in collecting International Relations data. The traditional approach in International Relations theory is to take the state as the key actor or fundamental unit. Hence, to answer questions about the relative frequency of wars between states, we clearly need to first identify the population of states that might enter into conflict with one another.

The literature on the state offers a number of conceptual definitions. An early influential example is Weber's (2004[1918]) definition of the state as "a relation of men dominating men" with a monopoly on the legitimate use of violence. More recently North (1981: 21) defines the state as "an organization with a comparative advantage in violence, extending over a geographic areas whose boundaries are determined by its power to tax constituents". Krasner's (1995/6) discussion of the possible changing and variable nature of the state highlights territoriality and autonomy as the defining characteristics of the Westphalian state.

Such efforts to develop conceptual definitions of the state rarely discuss how one might operationalize these defining characteristics in practice. The probably best-known effort to identify states in the international system empirically is the so-called Correlates of War list, first developed by Russett, Singer, and Small (1968). This list does not follow definitions of the state in emphasizing features such as territorial control and autonomy, but rather focuses on a set of criteria based on external recognition and a minimum population size, where the specific criteria for identifying states change over time. In brief, prior to 1920, the criteria for inclusion are i) recognition as a state by the UK and France *and* ii) a half a million minimum population threshold prior to 1920. After 1920, units are considered states if a) they are members in the League of Nations or the United Nations, *or* b) they exceed the half-million population threshold.

Whether this is a valid approach to identifying states depends upon the specific research question. Gleditsch and Ward (1999) criticize the content validity of the COW list criteria, since their operational definitions exclude a number of important autonomous actors with clear autonomy and territoriality in the 19th century when these were not recognized by either the UK or France. Moreover, the fact that different criteria are used over time means comparisons across time based on the COW list may be of questionable validity. Examples of questionable exclusions includes China and Iran, which were never colonized by any other state, as well as countries that were de facto autonomous in internal affairs but chose to let another state retain influence over foreign policy, such as the former British dominions Australia and Canada until after the Treaty of Versailles in 1919. Lemke (2002) notes that the COW list excludes many regionally important actors; In particular, many Latin American states appear in the COW list substantially later than their conventionally recognized dates of independence. More recently, the COW list has come to encompass a number of microscopic states, by virtue of UN membership, such as Palau (population 18,100), whose role as actors in world politics is somewhat questionable.

Gleditsch and Ward (1999) suggest alternative criteria that emphasize autonomy and territorial control rather than major power recognition, which may be more appropriate for researchers interested in making statements about independent states than the major power centered conception of the international system guiding the construction of the COW list. As can be seen from Figure 1, the Gleditsch and Ward data yield a substantially larger number of states in the 19th century than the COW list, and a lower number of states following the rush of many microstates to join the UN in the 1990s. Especially during the 19th century, the COW list leaves large parts of the globe excluded from population. As a consequence, it follows that any measure based on the number of states in the denominator could differ considerably depending on the definitions used.

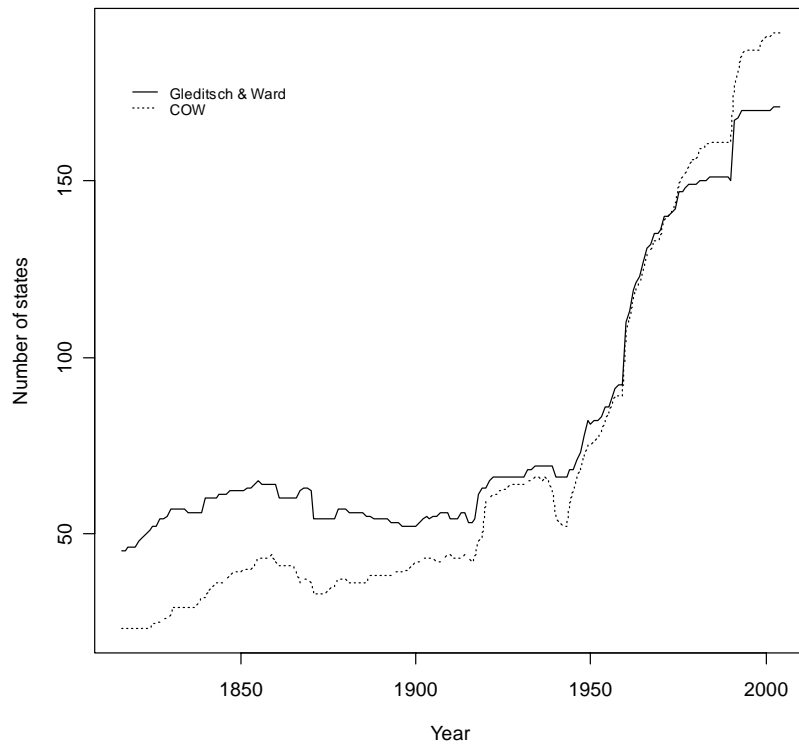


Figure 1: Comparison of population of states, Gleditsch & Ward and COW lists.

Bremer and Ghosn (2003) criticize Gleditsch and Ward for relying on subjective criteria to identify states. Using the language introduced above, the costs of seemingly more subjective criteria that may be difficult to apply in an intersubjective manner must be considered relative to the potential problems of content validity of alternative criteria that may appear more objective. External recognition can reflect political considerations as much as whether states have the characteristics that we expect states to have such as autonomy or territorial control. Objections from particular states can for example delay or prevent recognition by international organizations. This aside, even if the seemingly objective nature of the criteria might seem an attractive feature of the COW list, the stated rules are not actually sufficient to replicate the list. Gleditsch and Ward (1999) point out how Singer and Small excluded a

number of states that met their pre-1920 criteria and included others that did not, based on various ad-hoc decisions outside the explicit criteria. Ukraine and Belarus, for example, were independent UN members during the Cold War and signatories to the original UN charter, yet neither was considered a system member in the COW list.⁶

Identifying Conflict between States

In this section, we focus on how differences in definitions of conflicts used in various conflict data collection efforts will influence our inferences on trends in conflict and peace. As we have already mentioned, our decisions about what to consider a state will have obvious implications for our conclusions about conflicts between them. Gleditsch (2004b) identifies 22 additional wars between states in the Gleditsch and Ward list that satisfy the COW project's criteria for what constitutes an interstate war, but which were excluded from their data because at least one of the states involved was not recognized as a system member. However, since most efforts to collect data on conflict tend to use a particular definition of the population of states, it is generally difficult to ascertain how changing the criteria for states might influence our conclusions.

Before we can make any inferences about trends, we obviously need to define what we mean by conflict and peace. Here again, we have a rich conceptual literature offering many definitions (see, e.g., the extensive review in Most and Starr 1989). Boulding (1963: 5), for example, proposes that conflict can be defined “as a situation of competition ...[where] parties are aware of the incompatibility of their potential future position and ... [each] wishes to occupy a position that is incompatible with the position of the other.” The emphasis on

⁶ Gleditsch and Ward do not include these states as their independence from the Soviet Union is judged to be questionable (or largely fictitious). More generally, the COW list relies on similar subjective judgements supplementing or modifying the explicit operational criteria.

perceived conflict here allows us to exclude alleged forms of conflict that may not be understood by the actors, such as structural violence or certain forms of Marxist exploitation (Høivik and Galtung 1971; Roemer 1982). However, it is still far from clear how one would apply Boulding's definition in efforts to collect empirical data (see, e.g., Gleditsch 2002).

Despite the emphasis on incompatibilities as the defining characteristics in most conceptual definitions, most empirical conflict data collection efforts emphasize *manifest* forms of conflict or events involving the use of violence or perceived crisis. Incompatibilities may be enduring and take on the character of constant or regular features of relations – Spain, for example, continues to dispute UK sovereignty over Gibraltar, although it no longer acts in ways designed to enforce its claims. It would be nearly impossible to make an exhaustive catalogue over all such latent incompatibilities.⁷ By contrast, violent events and spectacular crises that are clearly departures from normal relations are much easier to observe and catalogue.

The first data collection efforts on conflict emphasized the use of violence as a distinguishing characteristic of events,⁸ and this approach was later adopted by the Correlates of War project (Singer and Small 1972). The COW project in essence defined wars to be events causing more than 1,000 casualties and provided data for the period following the Congress of Vienna in 1814–15. The COW war definition has the advantage of separating wars or large events from minor controversies or quibbles. Moreover, the core definition appears – perhaps deceptively – simple and easy to apply. Upon closer inspection, however, many ambiguities arise with regards to how one would identify wars from this definition.

⁷ The Issue Correlates of War project (see <http://www.icow.org>) has started collecting data on claims to territorial or environmental resources, but their work remains quite limited geographically and temporally given the scope of this task.

⁸ See in particular Richardson's (1960) *Statistics of Deadly Quarrels*.

Should one count deaths of active combatants or formal government soldiers, or should we include battledeaths by civilians? How do we determine if deaths were directly caused by conflict? How do we identify the start and end dates of a conflict? There has been considerable confusion over the specific operational criteria for identifying wars in the COW project, fueled in part by how the criteria have changed over time and doubt as to whether the criteria have been applied consistently over time (Gleditsch 2004b; Sambanis 2004). Moreover, although deaths in principle can be counted, determining whether the death threshold has been met involves relying on estimates of battledeaths, which in practice vary widely for many conflicts (see Lacina and Gleditsch 2005). The paucity of information on many parts of the world and time periods implies that the quality of data on casualties is likely to be very low for many violent incidents (see Shirkey and Weisiger 2007). Hence, despite the apparent objective nature of the criteria, the resulting list of wars encompasses many subjective judgments, and it would be difficult – if at all possible – to replicate the list without additional information on the sources and subjective estimates used.

Although a large number of studies have analyzed the incidence of war, many scholars have become concerned that looking only at the conflicts that eventually escalate to major wars may leave out much of the relevant conflictual situations and limit empirical variation that might help us better understand conflict and peace. Large wars are, fortunately, relatively uncommon events (Richardson 1948), and many argue that we can get a better understanding of under what conditions incompatibilities may escalate to violence if we take a more comprehensive look at crises and disputes between states from which war could arise.⁹

⁹ The fact that conflicts display a scale-invariant distribution where frequency is inversely proportional to magnitude could also be taken to suggest that small and large wars represent draws from the same underlying distribution where the events that give rise to large wars may not be inherently different from smaller conflicts (see, e.g., Cederman 2003; Richardson 1948). The issue of whether large wars could be understood within

Three alternative data collection efforts have tried to identify such broader sets of conflicts. The Uppsala Armed Conflict Data (ACD) project retains the focus on violence and identifies “armed conflicts”, where we see “a contested incompatibility that concerns government or territory or both where the use of armed force between two parties results in at least 25 battle-related deaths ...[of which one must be] the government of a state” (Gleditsch et al. 2002: 618-9). However, other data collection efforts have dropped the explicit emphasis on violence, and instead looked for criteria to help identify situations where violence may be likely. The Correlates of War project has produced a dataset on Militarized Interstate Disputes (MIDs), defined as “cases in which the threat, display or use of military force short of war by one member state is explicitly directed towards the government, official representatives, official forces, property, or territory of another state” (Jones, Bremer, and Singer 1996: 168). Brecher and Wilkenfeld (1997; 1988) have developed a dataset of international crises – the Interstate Crisis Behavior Data (ICB) – defined by a set of necessary criteria that together are sufficient: 1) “a threat to one or more basic values,” 2) “an awareness of finite time for response, and 3) “a heightened probability of involvement in military hostilities” (Brecher and Wilkenfeld 1997: 7).¹⁰

Although these data collection efforts certainly are very useful and commendable, the expanded scope of these efforts introduces additional potential problems to those that we have previously listed. If there is a problem in identifying major wars in information poor

general theories or require special explanations is debated in a special issue edited by Midlarsky (1990). Hegre and Sambanis (2006) find differences among robust predictors for conflicts based on different conflict thresholds.

¹⁰ Note that the ICB definition of crisis requires a judgment on whether or not the actors *perceived* a threat, finite time for response and heightened probability of military escalation. Hence, to determine what events constitute crises, coders must consider primary and secondary accounts of both the tangible actions transpired as well as how actors interpreted those actions.

environments as one goes back in time, then there is all the more reason to worry about whether we have sufficiently good historical data to identify such militarized interstate disputes and threats that may not entail any violence or casualties consistently in developing countries and in the 19th century. Gleditsch (2002: 000) notes that a disproportionate number of MIDs is reported for European states. Although this conceivably could be a “genuine” characteristic of the universe of militarized disputes, it seems more likely to arise in part as a result of more frequent reporting of events in this region and that many actual disputes that meet the definition simply go undetected in areas off the radar of international media. Moreover, the criteria for determining whether incidents are part of the “same” dispute or separate disputes in the MID data are complex. While wars such as the World Wars and the Mexican American War are considered one event with a single dispute ID, other contentious issues such as the Iranian threat to impose a blockade of the Strait of Hormuz give rise to a large number of disputes with distinct ID numbers.¹¹

Individuals in the COW project have often criticized the ICB data for relying on subjective evaluations of the perceptions of actors in identifying crises. However, it is highly

¹¹ The criteria for how incidents – the building blocks in the MID data – are aggregated into disputes are quite complex, and we refer to Jones et al. (1996 174-7) for the stated rules. In essence, the key criteria for determining whether individual incidents are part of an ongoing dispute hinge on a combination of whether incidents are seen as involving the “same” incompatibility, the time between evidence, whether there is evidence of coordination between parties, and whether an incident resulted in a formal treaty. Obviously, such decisions require considerable subjective judgment, and decisions are difficult to evaluate as the underlying incident data for the pre-1993 MID data are not publicly available (Glenn Palmer, personal communication). See also Fordham and Sarver (2001) for an insightful discussion of the MID criteria and possible problems for the study of US resort to force.

questionable whether one avoids the problem of subjective judgments by relying on secondary news report to identify disputes, as does the MID project, as these obviously also may reflect subjective perceptions or contentious views. Thompson (1995) similarly criticizes the content validity of rivalry measures based on the frequency of MIDs (see, e.g., Diehl and Goertz 2000), as these often include states that did not perceive one another as rivals and may omit rivals that did not record the required number of disputes.

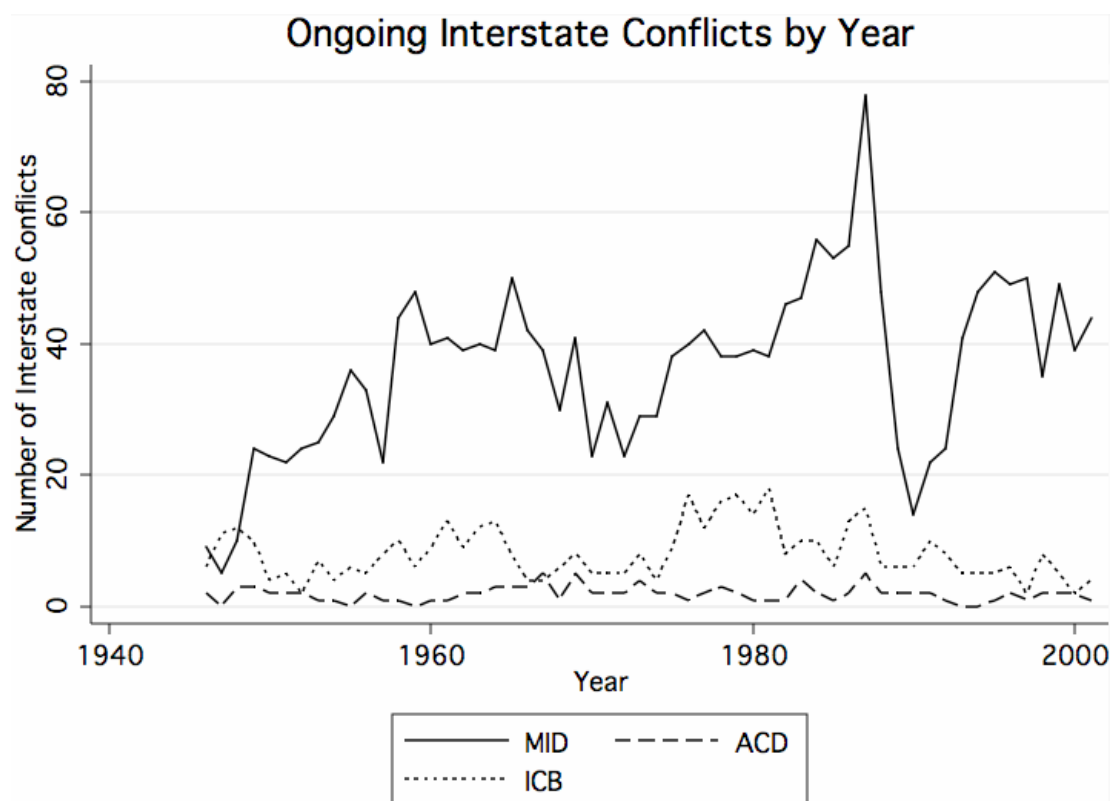


Figure 2: Number of ongoing interstate conflicts in the ACD, MID, and ICB data

Our brief overview of the definitions of the different conflict measures indicates that the different conflict data have quite different criteria for what constitutes a conflictual event. Each of these will in turn give quite different indications of the prevalence of conflict in the international system. Figure 2 displays the number of ongoing conflicts in each year, from 1946 to 2001, according to each of the datasets. As is evident, the MID data consistently

suggest far more conflictual events between states in each year than the other two data sources. Moreover, using the MID data as our guide we probably would not conclude that there was any decline in interstate war over time or a post-Cold War dip. Although the number of MIDs falls sharply at the end of the Cold War, this fall is preceded by an all-time high peak, and the number of MIDs quickly reverts back again to a level quite similar to the historical average. The ICB data suggest a much more restricted number of conflictual interstate events than the MID data. The ICB data do suggest a decline in conflict, but this here appears much earlier than the end of the Cold War and resembles a constant decline from the late 1980s, possibly as a result of improving relations between the Soviet Union in the West. When looking at interstate conflict only, the ACD data do not suggest much of a trend in Figure 2. However, this really reflects the sparsity of violent interstate armed conflict at any time during this period; hence the line hence appears almost entirely flat.¹²

These are all quite different conceptualizations of conflict, and as we have mentioned previously, there is no reason why different data necessarily should yield similar answers or trends. However, it is incumbent upon researchers to ensure that the conflict measure used actually corresponds to the concept of interest. Any threshold criteria will always to some extent be arbitrary and judgment is required in determining what counts as a particular type of conflict and what does not. For example, the “use of force category” in the MID data includes fishing disputes, and analysts should think carefully about whether a confrontation between the US and Canada set off by a fishing vessel on one side and the Coast Guard on the other is

¹² An analysis of the ACD data that includes intrastate actors reveals many more conflicts and variation over time.

as relevant in the study of crisis dynamics as rapidly escalating and potentially very damaging conflicts such as the one between India and Pakistan in 2001.¹³

A final problem in identifying conflict between states arises from its relationship to conflict within states. Civil wars, or conflict within states, is by far more common than conflicts between states (see Gleditsch et al. 2002; Holsti 1996). However, whereas many conflict data collection efforts impose typologies where there is a mutually exclusive distinction between *intra-* and *inter-*state wars, many conflictual events do not fit easily into these categories. Imposing such binary distinctions has led to numerous ambiguities in many data sources. In the Correlates of War data, for example, the civil war in Vietnam formally “ends” on 6 February 1965 when the US started the bombing of North Vietnam and the conflict was reclassified as an interstate war. And the Kashmir conflict has shifted back and forth between the civil and interstate war categories in different versions of the dataset, since a conflict by definition must be one of the two types.¹⁴ Similar issues arise in event-based data collections, and a large share of the recorded MIDs between states appear to originate out of civil conflicts rather than having interstate incompatibilities as their initial cause (see Gleditsch, Salehyan, and Schultz 2008). Although prior typologies and taxonomies can be helpful in data collection and theory building, these should not be allowed to become shackles that we strive to impose on our data if the observations refuse to behave according to our coding scheme. If a high share of disputes arise out of civil wars, then it seems questionable whether models focusing exclusively on state-to-state relations can realistically be expected to

¹³ Incidentally, if the Kargil conflict is a major war, then this would be the only case of direct war between two nuclear countries.

¹⁴ The Uppsala ACD data are an exception here, as they code conflicts by actors, so that one named conflict may include both state vs. state and state vs. non-state actor dyads. It is still possible to compute separate time-series for interstate and intrastate conflicts, but the starting-point is a joint list.

have high predictive capacity or include the relevant “issues” that may lead to conflict (see Diehl 1992; Gleditsch, Salehyan, and Schultz 2008).

Denominator effects in International Relations Research

In this section, we highlight how the answers that researchers reach may depend on what we call “denominator effects” in the construction of seemingly descriptive measures from the data.

We start by the issue of the relationship between the distribution of capabilities in the system and its sensitivity to the list of states in the system. Traditional international relations theory has emphasized the distribution of military capabilities as a key determinant for the prospects for conflict and peace. A central debate has been the relative merits of different forms of systemic polarity or balance of power, or the pacifying effects of power preponderance (e.g., Kaplan 1957; Waltz 1979). Hence, it is not surprising that initial empirical efforts in conflict studies sought to evaluate such systemic capability concentration arguments empirically. An early example here is Singer, Bremer, and Stuckey (1972), who examine the relationship between a measure of systemic concentration and the extent of war in the international system. Their proposed measure of systemic concentration is

$$CON = \sqrt{\frac{\sum s_i^2 - 1/n}{1 - 1/n}},$$
 where s_i indicates the proportion or share of capabilities held by a

state in a system, while conflict is measured in terms of the number of nation-months of war during each year. The results in Singer et al. (1972) suggested that war was less common the more dispersed capabilities were in the 19th century, which they saw as consistent with the idea of balancing favoring peace. However, greater concentration was associated with peace in the 20th century, supporting the notion that preponderance favors peace. Singer et al. (1972) argue, in a somewhat ad hoc fashion, that changes in the nature of diplomacy may account for

the shift between the two time periods, due to the increasing uncertainty resulting from a larger role for domestic politics in the 20th century.

The Singer et al. (1972) findings have been subject to a great deal of debate, and it is certainly possible to question the correspondence between their measure of power concentration and the concept of balance of power as well as the adequacy of their measures of conflict.¹⁵ Here, we wish to draw attention to how the differences in the definition of states across time can create problems for studies of this type. The CON measure is normalized by the number of states n and will be sensitive to changes in the number of states. This is due both to the presence of n itself in the formula for CON as well as from the fact that the proportional size s_i of a given amount capabilities that a state i holds must depend on the number of states in the system, since proportions are normalized so that $\sum s_i \equiv 1$. By implication, a fixed level of capabilities for two states could coincide with changing systemic proportions simply because more states appear in the system and add to the denominator. As we mentioned above, measures based on shares of states will be difficult to compare across shifts in the COW criteria identifying the population of states. Although Singer et al. (1972) look only at major powers, the amount of warfare in the system will obviously also be influenced by what is left in and out of the denominator or population. Hence, we should be very cautious in drawing strong conclusions about differences over time being due to changing features rather than possible artifacts of our measures and definitions. In this instance, it is difficult to ascertain the specific consequences of the restrictive definition of states, since we do not have supplementary data on military capabilities for states not included in the COW list. However, in principle this could be investigated empirically.

¹⁵ Vasquez (1993) provides a thoughtful discussion of criticism and subsequent reevaluation.

We turn instead to another area where it is much easier to demonstrate denominator effects, the question of the prevalence of democracy in the world. Huntington (1991) and others have suggested a wave-like pattern in the expansion and contraction of democracy in the world. Some international relations researchers have examined how such changes in the share of democracies either globally or regionally is associated with changes in conflict and peace (see, e.g., Crescenzi and Enterline 1999; Kadera, Crescenzi, and Shannon 2003). Figure 3 shows a plot of the proportion of independent states with a score of 6 or above on the POLITY democracy scale (Jagers and Gurr 1995). This does indeed suggest a pattern of three waves of democracy, followed by two waves of reversals.

Figure 3 is based on using the number of states in the denominator. We already know from Figure 1 that the number of states in the international system is not constant, but has increased rapidly with the decline of larger empires and the process of emancipation of former colonies. If the denominator n is not constant, then the share of democracies could decline, even if we have no actual reversals from democracy to autocracy. More problematically, the types of states that have been added to the system with de-colonialization are likely to be very different from long-established states, and typically have much lower levels of economic development or human capital. For example, in studies of economic openness over time, Alesina, Spolaore, and Wacziarg (2003) show that normalizing by the changing number of states suggests a declining degree of interconnectedness. However, this is due to changes in the type of state in the population rather than a decline in trade between existing states. Similarly, few previous colonies start out as democracies, but are likely to emerge with autocratic institutions. Doorenspleet (2000) holds that much of the apparent evidence for waves of democratization is due to changes in the system rather than changes between democracy and autocracy in state institutions.

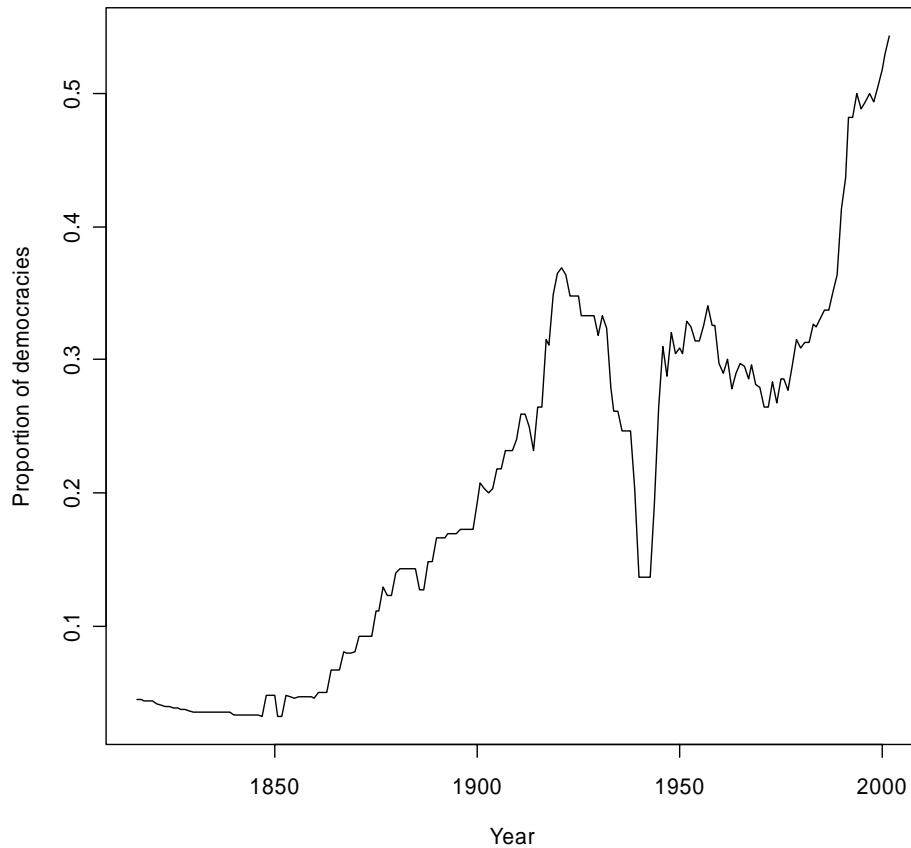


Figure 3: The proportion of democratic states

In Figure 3 all states have equal weight. However, some states carry more weight in international politics than others. For example, the regime type of India and China is much more significant both for conflict and global demonstration effects than the status of Tonga and Palau. One possibility would be to weight countries by some measure of their relative share, for example population, as shown in Figure 4. This suggests a much higher level for the second wave than the proportion of states in Figure 3, reflecting the fact that India and other large states are accorded more weight than smaller states. However, Figure 4 does not indicate a dramatically different picture in terms of the waves of democracy.

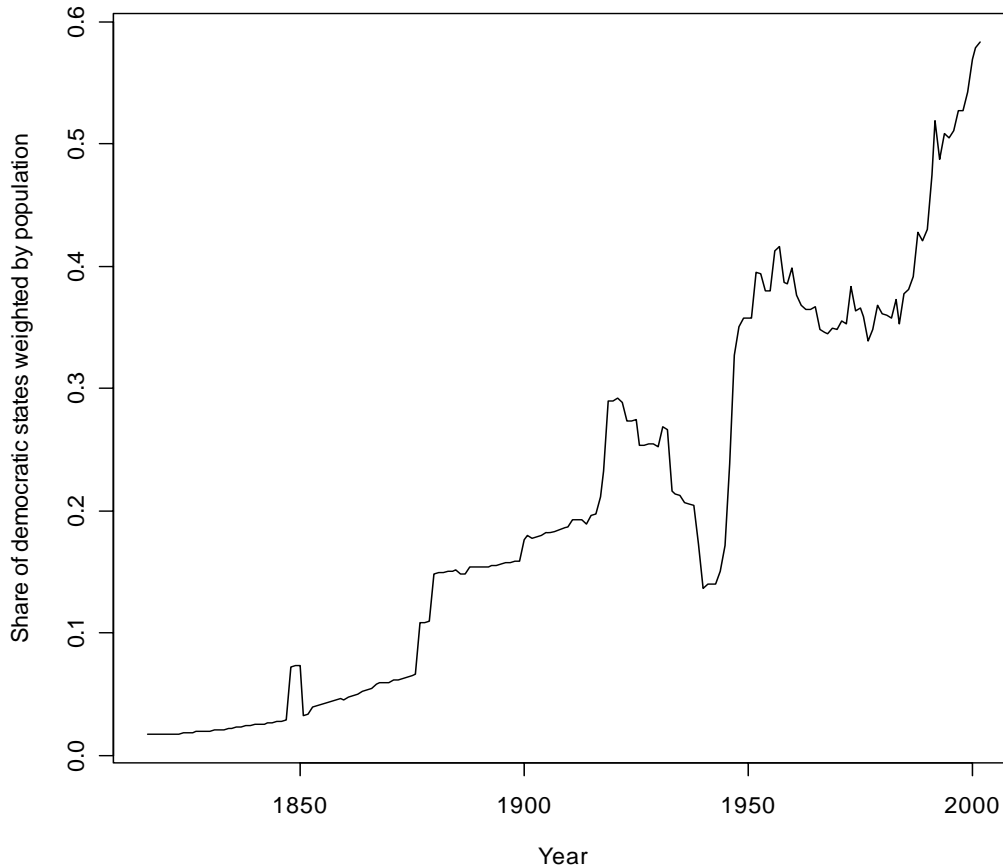


Figure 4: Share of democracies, weighting states by population

A more subtle issue that has received little attention so far is that if we are interested in the extent to which the world is democratic, then it is not obvious that we want to limit the denominator to the part living in independent states. The principle of the right of national self-determination is very much a post-1945 phenomenon, and the further back in time we go, a substantial share of the world's population will be in dependent areas or colonies, under the control of other states. As we have shown above, when the denominator increases as former colonies become independent, we will get an increase in the number of non-democratic independent states. However, since the populations in these areas were previously living under non-democratic colonial rule, it is misleading to argue that autocracy is becoming

widespread simply because these areas now have independent autocratic states as opposed to colonial non-democratic administration.

Unlike many other forms of data that are only collected for state-like units, it is possible to get estimates for world population (or estimates by region), including individuals living in colonies and dependent areas. Hence, we can get a better measure of the extent of democracy in the world by looking at the share of population living in democratic states over the total global population. Using expanded population data collected by (Gleditsch 2004a), Figure 5 suggests that there was a real setback for democracy in the period leading up to World War II. However, this is the only major period of democratic setback over the two last centuries. Even the decolonization period emerges as a static period, where democracy does not expand, rather than a retrenchment. According to these estimates, it is only around the turn of the millennium that a majority of the world's population live under democratic institutions.

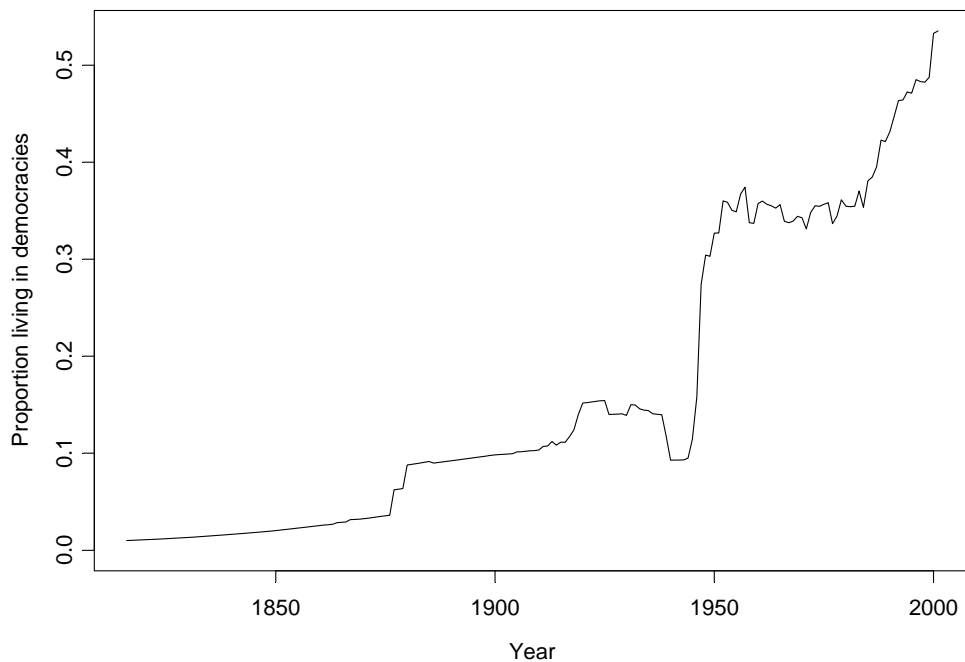


Figure 5: Share of world population living under democratic rule

The large denominator effects shown here imply quite different conclusions about the frequency and changes in democratic institutions. Our point to emphasize how conclusions in analyses on war and peace can be influenced by denominator effects or non-random missing data rather than arguing that one measure is inherently superior to another. The main lesson is that scholars should carefully specify empirical measures that actually correspond to their concept of interest rather than uncritically use what is most readily available.

Recommendations for future research

Our discussion so far has generally had a pessimistic tone, pointing to potential problems arising in data collection. This is not to be construed as an argument for abandoning empirical work and reverting to arm-chair theorizing. Data collection in international relations has indisputably facilitated a great deal of valuable systematic research. In numerous instances widely shared beliefs about alleged trends have been refuted by systematic data. Prominent examples here include the alleged more dangerous nature of the post-Cold War world, as well as the extreme fears over the alleged global erosion of democracy expressed in the 1970s and 1980s (e.g., Revel 1984; see also the interesting review in Mueller 1994: 362). Despite these successes, greater attention to problems in data collection and interpretation is important in its own right, and essential for advancing research on conflict and peace. In this section, we try to provide a more positive contribution with some practical recommendations on how data collection efforts ought to proceed.

First, the distinction between subjective and objective measures often is overstated. Leaving aside the question of content validity, most seemingly objective criteria (e.g., numbers of people killed) tend ultimately to rely on various subjective and potentially controversial judgments. The more serious problem plaguing many international relations data

collection efforts is that so little documentation is available detailing the specific judgments and sources going into the production of the data. The Correlates of War (COW) war data, for example, provide no documentation on the specific sources used to determine whether conflicts meet their battledeath criteria and why codings change, beyond some discussion of how their original data were compiled from existing data sources (e.g., Singer and Small 1972; Small and Singer 1982). Likewise, very little information has been available on coding of the MID disputes data, making it difficult for users to figure out what the events included actually may refer to, as well as what sources were used in coding the event.¹⁶ By contrast, although the ICB data may be criticized for relying on subjective judgments on perceptions, the ICB project has been quite explicit in documenting the rationale for their classifications (see e.g., the extensive case summaries in Brecher and Wilkenfeld 1997). Information on decisions and sources/origins of coding allows users to inspect these for individual observations in order to determine whether these are relevant for their particular research question or not.¹⁷

Second, data collection efforts should be much more explicit about the uncertainty in classifications. Recent data collection efforts by Lacina et al. (2006) and Valentino (2004), for example, provide a range of estimates for casualty numbers, which are very helpful in providing a sense of the uncertainty surrounding some of even the best guesses. Morrow (2007) similarly codes not only the degree of compliance with laws of war, but also the

¹⁶ Effort to identify additional information about the events in the MID data by Schultz reveal that it was difficult to find information about many of the events in the MID data. In some instances, the only source for the militarized events appears to be statements from the alleged target of the action, with the initiating state denying that the events has taken place (see Gleditsch et al 2008: 26). This hardly seems consistent with the definition provided by Jones et al. (1996) and the emphasis on explicit action.

¹⁷ Other examples of data with extensive source documentation include Lacina et al. (2006) on battledeaths and Vanhanen's (2000) democracy data.

quality of the data and the clarity of non-compliance. By incorporating uncertainty in projects, the collector allows future users to understand when the indicators involve more subjective judgment and help to flag decisions that other researchers may disagree with or exclude from their own analyses.

Third, expert opinion surveys can help to limit the impact that a single person's judgment has on the final resulting data. An early example of an expert-based survey is Goldstein's (1992) development of a cooperation-conflict index based on the WEIS event data scores. To create the scale, Goldstein had to assign values related to the level of cooperation or conflict involved in such events as "deny negotiations" and "promise military support." Instead of relying exclusively on his own best guesses, Goldstein surveyed international relations experts and had them assign values to each of the possible events. Goldstein then found the means and standard deviations around each of the assigned values and used these to construct the conflict-cooperation scale for the categories. In this way, Goldstein was able to reduce the influence that a single coder's biases can have on the resulting classifications, and the reported measures of uncertainty allow the user of the data to incorporate the uncertainty around these values. More recently, Baum and Groeling (forthcoming) polled international security experts to create an estimate for the degree to which the US was succeeding or failing in Iraq during various time periods. Cederman, Girardin, and Wimmer (2006) have conducted an ethnic survey of the political status of different ethnic groups using an internet-based platform that records the individual users rankings and the full revision history of the data. This approach provides an excellent way of making the coding transparent and allowing interested readers access to decisions that go into the data.

Finally, although much of data collection efforts in international relations have emphasized collecting data for the entire population or system, in many instances a targeted,

smaller data collection strategy might be a more cost-effective way to expand our knowledge (e.g., King and Zeng 2000). Owing to the predominance in systemic arguments in the discipline at the outset of the behavioral revolution, it is not surprising that early data-collection efforts sought to classify information for the international system at large. But as our discipline has increasingly turned to evaluating dyadic propositions, scholars now attempt explanations based on variation in dyadic rather than systemic characteristics. Elementary statistics courses are designed to teach us that making generalizations to a population does not require that we have data on the full population, and that much information can be gleaned from a random sample. If data are difficult to collect for a full population, then we will often be much better off trying to track down good information on a smaller sample while ensuring that we minimize bias in the information collected.

References

- Alesina, Alberto, Enrico Spolaore, and Romain Wacziarg. 2003. "Trade, Growth and the Size of Countries." In *Handbook of Economic Growth*, 1499-1542, edited by Philippe Aghion and Steven N. Durlauf. New York: North Holland.
- Bollen, Kenneth A. 1990. "Political Democracy: Conceptual and Measurement Traps". *Comparative Studies in International Development* 25:7-24.
- Boulding, Kenneth E. 1963. *Conflict and Defense: A General Theory*. New York: Harper and Row.
- Brecher, Michael, and Jonathan Wilkenfeld. 1997. *A Study of Crisis*. Ann Arbor, MI: University of Michigan Press.
- Brecher, Michael, Jonathan Wilkenfeld, and Sheila A. Moser. 1988. *Crises in the twentieth century: Handbook of international crises*. New York: Pergamon.
- Bremer, Stuart A., and Faten Ghosn. 2003. "Defining States: Reconsiderations and Recommendations". *Conflict Management and Peace Science*. 20 (1):21-41.
- Campbell, Donald T., and Julian C. Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.
- Cederman, Lars-Erik. 2003. "Modeling the Size of Wars: From Billiard Balls to Sandpiles". *American Political Science Review* 97 (1):135-150.
- Cederman, Lars-Erik, Luc Girardin, and Andreas Wimmer. 2006. Getting Ethnicity Right: An Expert Survey on Power Distributions among Ethnic Groups. Typescript, Center for International Studies, ETH-Zurich.
- Clauset, Aaron, Maxwell Young, and Kristian Skrede Gleditsch. 2007. "On the Frequency of Severe Terrorist Events". *Journal of Conflict Resolution* 51 (1):1-31.
- Crescenzi, Mark J. C., and Andrew J. Enterline. 1999. "Ripples from the Waves? A Systemic, Time-Series Analysis of Democracy, Democratization, and War". *Journal of Peace Research* 36 (1):75-94.
- Diehl, Paul F. 1992. "What are they Fighting for? The Importance of Issues in International Conflict Research". *Journal of Peace Research* 29:333-44.
- Diehl, Paul F., and Gary Goertz. 2000. *War and Peace in International Rivalry*. Ann Arbor, MI: University of Michigan Press.
- Dixon, William J., and Terry Boswell. 1996. "Dependency, Disarticulation, and Denominator Effects: Another Look at Foreign Capital Penetration". *American Journal of Sociology* 102 (2):543-562.
- Doorenspleet, Renske. 2000. "Reassessing the Three Waves of Democratization". *World Politics* 52 (3):384-406.
- Firebaugh, Glenn. 1992. "Growth Effects of Foreign and Domestic Investment". *American Journal of Sociology* 98 (1):105-30.
- Fordham, Benjamin O., and Christopher C. Sarver. 2001. "Militarized Interstate Disputes and United States Uses of Force". *International Studies Quarterly* 45 (3):455-466.
- Gleditsch, Kristian S., and Michael D. Ward. 1999. "A Revised List of Independent States Since 1816". *International Interactions* 25 (4):393-413.
- Gleditsch, Kristian Skrede. 2002. *All International Politics is Local: The Diffusion of Conflict, Integration, and Democratization*. Ann Arbor, MI: University of Michigan Press.
- Gleditsch, Kristian Skrede. 2004a. Expanded Population Data v1.0. Department of Political Science, University of California, San Diego.
- Gleditsch, Kristian Skrede. 2004b. "A Revised List of Wars Between and Within Independent States, 1816-2002". *International Interactions* 30 (4):231-262.
- Gleditsch, Kristian Skrede, Idean Salehyan, and Kenneth Schultz. 2008. "Fighting at Home, Fighting Abroad: How Civil Wars Lead to Interstate Disputes". *Journal of Conflict Resolution* tba (tba):forthcoming.

- Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand. 2002. "Armed Conflict 1946-2001: A New Dataset". *Journal of Peace Research* 39 (5):615–637.
- Goldstein, Joshua S. 2002. The Worldwide Lull in War. *Christian Science Monitor*, 14 May 2002, 9.
- Gurr, Ted Robert. 2000. "Ethnic Warfare on the Wane". *Foreign Affairs* 79 (3):52–64.
- Hegre, Håvard, and Nicholas Sambanis. 2006. "Sensitivity Analysis of Empirical Results on the Causes of Civil War Onset". *Journal of Conflict Resolution* 50 (4):508-535.
- Hewitt, Joseph, Jonathan Wilkenfeld, and Ted R. Gurr. 2007. *Peace and Conflict 2008*. Boulder, CO: Paradigm.
- Høivik, Tord, and Johan V. Galtung. 1971. "Structural Violence: A Note on Operationalization". *Journal of Peace Research* 7:73-76.
- Holsti, Kalevi J. 1996. *The State, War, and the State of War*. New York: Cambridge University Press.
- Human Security Report. 2005. *Human Security Report 2005: War and Peace in the 21st Century*. Oxford: Oxford University Press.
- Huntington, Samuel P. 1991. *The Third Wave: Democratization in the Late Twentieth Century*. Norman, OK: Oklahoma University Press.
- Jagers, Keith, and Ted R. Gurr. 1995. "Tracking Democracy's Third Wave with the Polity III Data". *Journal of Peace Research* 32 (4):469-482.
- Jones, Daniel M., Stuart A. Bremer, and J. David Singer. 1996. "Militarized Interstate Disputes, 1816-1992: Rationale, Coding Rules, and Empirical Applications". *Conflict Management and Peace Science* 15 (2):163-213.
- Kadera, Kelly M., Mark J. C. Crescenzi, and Megan L. Shannon. 2003. "Democratic Survival, Peace, and War in the International System". *American Journal of Political Science* 47 (2):234-247.
- Kaldor, Mary. 2006. *New Wars and Old Wars: Organised Violence in a Global Era, 2nd edition*. Cambridge: Polity.
- Kalyvas, Stathis N. 2001. "'New' and 'Old' Civil Wars: A Valid Distinction?" *World Politics* 54 (1):99-118.
- Kaplan, Fred. 2006. What "Peace Epidemic"? Don't pop the champagne corks just yet, the evidence isn't quite there. Online magazine, posted 25 January, <http://www.slate.com/id/2134846/?nav=tap3>.
- Kaplan, Morton A. 1957. *System and Process in International Politics*. New York: Wiley.
- King, Gary, Christopher J.L Murray, Joshua A. Salomon, and Ajay Tandon. 2003. "Enhancing the Validity and Cross-cultural Comparability of Measurement in Survey Research". *American Political Science Review* 97 (4):567-584.
- King, Gary, and Langche Zeng. 2000. "Logistic Regression in Rare Events Data". *Political Analysis* 9 (2):137-163.
- Krasner, Stephen D. 1995/6. "Compromising Westphalia". *International Security* 20 (3):115-151.
- Lacina, Bethany Ann, Nils Petter Gleditsch, and Bruce M. Russett. 2006. "The Declining Risk of Death in Battle". *International Studies Quarterly* 50 (3):673–680.
- Lacina, Bethany, and Nils Petter Gleditsch. 2005. "Monitoring Trends in Global Combat: A New Dataset of Battle Deaths". *European Journal of Population Studies* 21 (2-3):145-166.
- Lemke, Douglas. 2002. *Regions of War and Peace*. Cambridge: Cambridge University Press.
- Mack, Andrew. 2002. "Civil War: Academic Research and the Policy Community". *Journal of Peace Research* 39 (5):515–525.
- Mearsheimer, John J. 1990. "Back to the Future: Instability in Europe after the Cold War". *International Security* 15:5-56.
- Midlarsky, Manus I. 1990. "Big Wars, Little Wars - A single Theory?" *International Interactions* 16 (2):tba.
- Morrow, James D. 2007. "When Do States Follow the Laws of War?" *American Political Science Review* 101 (3):559-572.
- Most, Benjamin A., and Harvey Starr. 1989. *Inquiry, Logic, and International Politics*. Columbia, SC: University of South Carolina Press.
- Mueller, John. 1994. "The Catastrophe Quota: Trouble After the Cold War". *Journal of Conflict Resolution* 38 (3):355-375.
- Mueller, John. 2004. *The Remnants of War*. Ithaca, NY: Cornell University Press.

- North, Douglas C. 1981. *Structure and Change in Economic History*. New York: Norton.
- Revel, Jean-Francois. 1984. *How Democracies Perish*. New York: Doubleday.
- Richardson, Lewis F. 1960. *Statistics of Deadly Quarrels*. Chicago, Pittsburgh, PA: Quadrangle/Boxwood.
- Richardson, Lewis F. 1948. "Variation of the Frequency of Fatal Quarrels with Magnitude". *Journal of the American Statistical Association* 43 (244):523-546.
- Roemer, John E. 1982. *A General Theory of Exploitation and Class*. Cambridge, MA: Harvard University Press.
- Rubin, Donald. 1976. "Inference and Missing Data". *Biometrika* 63 (3):581-592.
- Rummel, Rudolph J. 1995. *Death by Government*. New Brunswick, NJ: Transaction Publishers.
- Russett, Bruce M., J. David Singer, and Melvin Small. 1968. "National Political Units in the Twentieth Century: A Standardized List". *American Political Science Review* 62 (3):932-951.
- Sambanis, Nicholas. 2004. "What is a Civil War? Conceptual and Empirical Complexities of an Operational Definition". *Journal of Conflict Resolution* 48 (6):814-858.
- Sarkees, Meredith Reid, Frank Whelon Wayman, and J. David Singer. 2003. "Inter-State, Intra-State, and Extra-State Wars: A Comprehensive Look at Their Distribution over Time, 1816–1997". *International Studies Quarterly* 47:49–70.
- Shirkey, Zachary C., and Alex Weisiger. 2007. An Annotated Bibliography for the Correlates of War Interstate Wars Database1.
<http://www.correlatesofwar.org/cow2%20data/WarData/InterState/Shirkey%20&%20Weisiger%20War%20Bibliography.pd>.
- Singer, J. David, and Melvin Small. 1972. *The Wages of War, 1816-1965: A Statistical Handbook*. New York: Wiley.
- Singer, J. David, Stuart Bremer, and John Stuckey. 1972. "Capability Distribution, Uncertainty, and Major Power War". In *Peace, War, and Numbers*, 19-48, edited by Bruce M. Russett. Beverly Hills, CA: Sage.
- Small, Melvin, and J. David Singer. 1982. *Resort to Arms: International and Civil Wars, 1816-1980*. Beverly Hills, CA: Sage.
- Summers, Robert, and Alan Heston. 1988. "A New Set of International Comparisons of Real Product and Price Levels for 130 Countries, 1950-1985". *Review of Income and Wealth* 34:358-21.
- Thompson, William R. 1995. "Principal Rivalries". *Journal of Conflict Resolution* 39:195-223.
- Valentino, Benjamin. 2004. *Final Solutions: Mass Killing and Genocide in the Twentieth Century*. Ithaca, NY: Cornell University Press.
- Vanhanen, Tatu. 1990. *The Process of Democratization: A Comparative Study of 147 States, 1980-88*. New York: Crane Russak.
- Vanhanen, Tatu. 2000. "A New Dataset for Measuring Democracy, 1810-1998". *Journal of Peace Research* 37 (2):251-265.
- Vasquez, John A. 1993. *The War Puzzle*. Cambridge, New York: Cambridge University Press.
- Wallensteen, Peter, and Margareta Sollenberg. 1995. "After the Cold War: Emerging Patterns of Armed Conflict 1989-94". *Journal of Peace Research* 32 (3):345-360.
- Waltz, Kenneth N. 1979. *Theory of International Politics*. Reading, MA: Addison-Wesley.
- Weber, Max. 2004[1918]. *The Vocation Lectures: 'Science as Vocation', 'Politics as Vocation'*. Indianapolis, IN: Hackett.
- Wilson III, Ernest J., and Ted Robert Gurr. 1999. Fewer Nations are Making War. *Los Angeles Times*, 22 August.