Online Appendix for
Aiding War: Foreign Aid and the Intensity of Violent Armed Conflict

Michael G. Findley, Department of Government, UT Austin
Daniel Strandow, Department of Peace and Conflict, Uppsala University
Joseph K. Young, School of Public Affairs, American University
Olgahan Cat, Department of Government, UT Austin

5 December 2020

Descriptive Statistics

Table 1. Independent (treatment) variables

Purpose	Name	Description
Hypothesis testing	Funding per Location	Coded 1 if the value (constant USD) per aid location is over the mean, 676,274
Robustness	Funding per Location, varying the threshold	Six thresholds to either side of the mean. Deviating from the mean with -0.15 to +0.15 standard deviations, in 0.05 increments

Table 2. The origins of the dependent variables from the Uppsala Conflict Data Program measurements

Violence Intensity	Casualties from	Measurement
Military Fatalities	Government, Rebel, or Militia Troops	Side A and Side B Deaths; Unknown Deaths
Civilian Fatalities	Civilians caught in Crossfire, Indiscriminately or Selectively Targeted	Civilian Deaths from State-based, Non-state, or One-sided Violence

Table 3. Preparing the dependent variables for analyses

Name	Description
Military Fatalities Log	T+1, sum of best estimates of all fatalities minus civilian deaths (log-10 of value+1)
Civilian Fatalities Log	T+1, sum of all civilians killed by either side (log-10 of value+1)
Total Fatalities Log	T+1, sum of best estimates of all fatalities (log-10 of best estimate+1)

Table 4. Time lags and trends control variables

Name	Description
1989	The first covariate year in the dataset, coded 1 if 1989
…	All years in between 1989 and 2008
2008	The last covariate year in the dataset, coded 1 if 2008
Funding Concentration, t-1	T-1 version of Funding Concentration, coded 1 if over 417,245 USD
Funding Concentration, t-1, varying the threshold	Six thresholds to either side of the mean of T-1 Funding Concentration. Deviating from the mean with -0.15

	to +0.15 standard deviations, in 0.05 increments
Funding per Area, t-1	T-1 version of Funding per Area
Total Funding, t-1	T-1 version of Total Aid
Civilian Fatalities, lag1	First lag, sum of civilian deaths (log-10 of fatalities +1)
Military Fatalities, lag1	First lag, sum of best estimates of all deaths excluding civilian deaths (log-10 of fatalities+1)
Total Fatalities, lag1	First lag, sum of best estimates of all deaths, log-10 of value+1
Civilian Fatalities, lag2	Second lag, sum of civilian deaths (log-10 of fatalities +1)
Military Fatalities, lag2	Second lag, sum of best estimates of all deaths excluding civilian deaths (log-10 of fatalities+1)
Total Fatalities lag2	Second lag, sum of best estimates of all deaths, log-10 of value+1

Table 5. Attacks, control, and spatial diffusion of attacks

Name	**Description**
Greater	*A* Preponderance in Control over Population.
Battleground	Coded 1 if *A* had a difference in population

Control	affected by control > 73580 (twice the average difference)
Greater Battleground Control, Alternative	For robustness. *A* more Control Counts. Coded 1 if *A* asserted control over more territory than *B* during current year and area
A is Challenger	Whether *A* is a challenger
Multiple Opponents	Coded 1 if multiple opponents in area
Attacks by A	Sum of all points attacked by *a* in administrative division
A over Peer Attacks	Dichotomous. Coded 1 if current area has as great, or greater, number of attacks by *A* than all other areas within the country that party *A* operates in
Population near Violence	Mean size of populations at battle locations

Table 6. Resource value control variables

Name	**Description**
Petro Locations	Number of petro locations within administrative division
Diamond	Number of diamond locations within

Locations	administrative division
Population Density	Population density
Rainfall	Rainfall in percentages
Agriculture	Agriculture land (land used for crops or pastures) coded in the following increments: 14%, 16%, 20%, 50%, 70%
Most Petro	Dichotomous. Coded 1 if current area has *greater* number of petro locations than all other areas within the country that actor a operates in
Most Diamonds	Dichotomous. Coded 1 if current area has *greater* number of diamond locations than all other areas within the country that actor a operates in
Most Agriculture	Dichotomous. Coded 1 if current area has as great, or greater, crops or pastures area percentage than the neighborhood max

Table 7. Geography Control Variables

Name	**Description**
Capital	Dichotomous, coded 1 if national capital in area
Mountainous	Real values of minimum elevation in meters
Forested	Percentages of forest cover
Most Mountainous	Dichotomous. Coded 1 if current area has as great, or greater, elevation than the neighborhood max

Most Forested	Dichotomous. Coded 1 if current area has as great, or greater, forest percentage than the neighborhood max
Area Size	Area in square kilometers
Greatest Area	Dichotomous. Coded 1 if current area has greater square kilometer area than all other areas within the country that actor a operates in

Table 8. Descriptive statistics of the independent variable

	Treatment = 1	**Mean**	**Standard deviation**
Aid per location	367	0.15	0.36
N=2378			

Table 9. Descriptive statistics of the dependent variables

	Mean	**Standard deviation**
Military fatalities, log	0.64	0.93
Civilian fatalities, log	0.34	0.66
Total fatalities, log	0.76	0.98
N=2378		

Additional Robustness Check

As the final robustness check, we use propensity score matching to address selection effects. When using exact matching, a subject under treatment is paired with a control subject if the two share exactly the same value on all covariates except the key independent variable. Propensity score matching instead pairs subjects based on how likely they are to receive treatment, which occurs when their propensity scores are similar (Rosenbaum & Rubin, 1983; Sekhon, 2009). There are a number of model specifications that can be used to estimate the propensity score. Our treatment variables are dichotomous, and we use a logit specification (Caliendo & Kopeinig, 2008). The procedure is less dependent on model assumptions than equivalent procedures that achieve as-if random assignment in regression models (Rosenbaum & Rubin, 1983, pp. 48–49). This benefit does, however, come at the expense of using a coarse dichotomous independent variable. For robustness, we vary the threshold used to generate this variable and show all possible threshold ranges for which the effect holds. Comparing the effect of treatment and control observations – within matched pairs – on the outcome makes it possible to estimate the effect of aid concentration on military fatalities. Propensity score matching can be used in combination with regression for more accurate results, though introducing some dependence on regression model assumptions (Ho, Imai, King, & Stuart, 2007, pp. 200, 209–211). Rather than doing post-matching regression we trade some accuracy for fewer model assumptions, simplicity of analysis, and high transparency, by calculating the average treatment effect (ATE), unlike we did with Matching Frontier.[1]

The ATE gives the difference in expected values of outcomes between observations of treatment and observations of control (Morgan & Winship, 2007, pp. 36–37). Because we

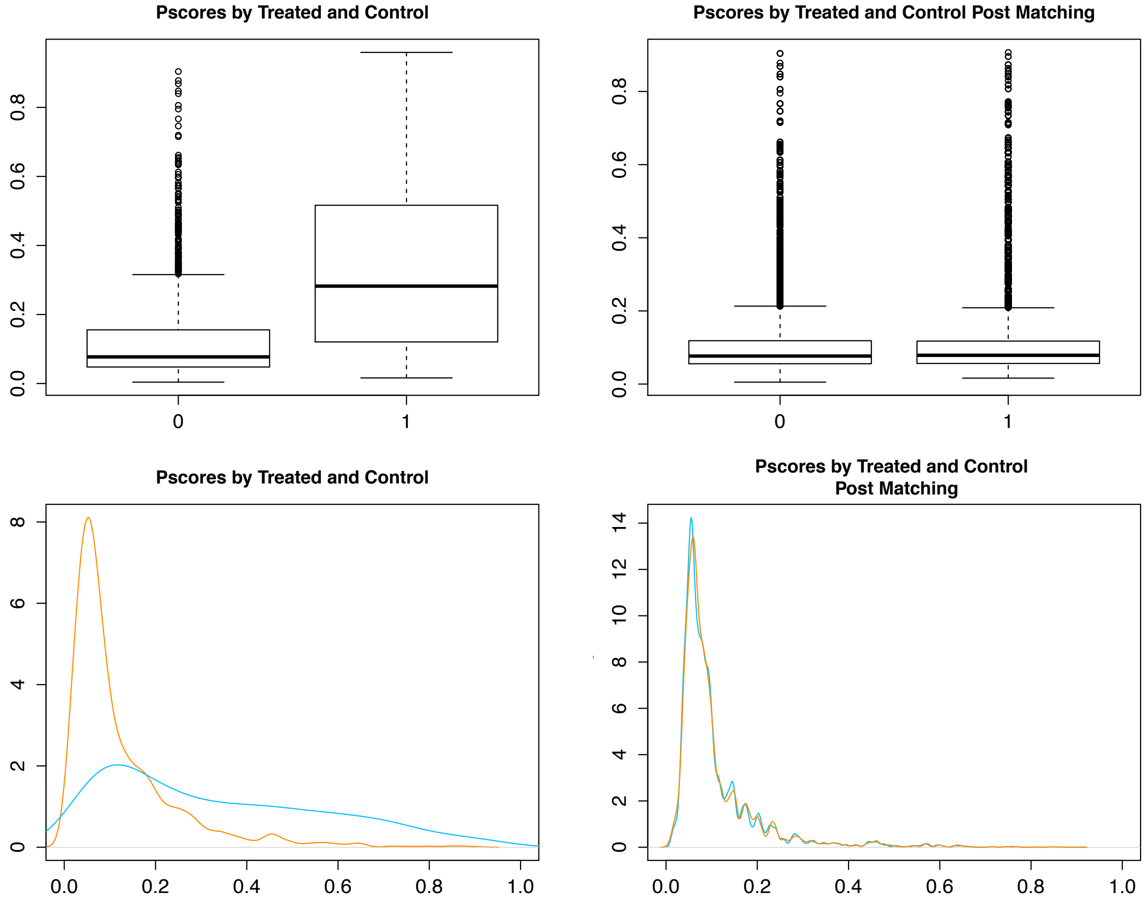
[1] The decreased accuracy results from the remaining "imbalance in the matched sample [that is] is strictly unrelated to the treatment [..], or [that] has no effect on the outcome" (Ho, Imai, King, & Stuart, 2007, p. 213).

are interested in the average effect over both treatment and control, control observations are matched to treatment observations and vice versa (cross-matching; compared to the average treatment effect of the treated in Ho et al., 2007, p. 216).

The research design we have specified results in matched pairs of treatment and control observations that are more likely to, for instance, contain warring parties within the same area, the same year, the same type, or with the same amount of opponents. This means that on occasion an actor could be compared to itself at a later date, or possibly to its current opponent. We consider that this design is appropriate in order to correct for the time and space dependent effects that unobserved covariates might have on the ATE.

As discussed above, the treatment variable funding concentration, is operationalized as *Funding per location*. Figure 4 shows propensity scores before matching to the left and the propensity scores of observations that remain after matching (post matching) to the right.

Figure 1: Pre- and post-matching of Funding per location, for military deaths



The graphs in the upper row show how the average propensity score (y-axis), and the spread in scores, varies between treatment and control observations (x-axis). The propensity to receive treatment is close to 0.3 for the treated observations. The lower row makes a similar point by showing the cumulative propensity score on the y-axis and the propensity score on the x-axis. The curve representing treated observations is colored blue and in the figure to the left it is the flatter of the two. The cumulative propensity score essentially adds together the number of observations of a certain propensity score so that it is possible to visualize which scores that are more common. The lower figure echoes the box-plot in showing that the propensity scores for the treated observations are more spread out than those of the untreated observations. After matching, the distributions of treatment and control observations are well balanced.

Table 10 displays the results of the post-matching difference tests. In considering the main effects, we find that our hypothesis is supported. Specifically, if aid funding is expected to be concentrated rather than diffused, the short-term *military* fatalities increase. We find no effect on civilian deaths suggesting that in already violent areas more concentrated funding tends to shift the mode of warfare between armed groups, and not the intensity in one-sided violence. Aid per location is associated with an increase in total deaths (civilian plus military deaths) but we would expect this result to be driven by the impact of funding concentration on military deaths. Control variables that are included in a model specification are indicated with check marks. Calculating treatment effects of control variables is irrelevant since they are most likely not as-if randomly assigned.

Table 10: Average Treatment Effect of aid value per location

Variable	Total Deaths	Military Deaths	Civilian Deaths
Funding per location	0.27***	0.31***	0.06
	(0.099)	(0.097)	(0.067)
Two-tailed p-value	0.006	0.001	0.395
Greater battleground control	✔	✔	✔
Number of petro locations	✔	✔	✔
Number of diamond locations	✔	✔	✔

Number of attacks committed by party A	✓	✓	✓
If A is challenger	✓	✓	✓
If A has multiple opponents	✓	✓	✓
Average population near battlegrounds	✓	✓	✓
Capital	✓	✓	✓
Area size	✓	✓	✓
Population density	✓	✓	✓
Precipitation	✓	✓	✓
Minimum elevation	✓	✓	✓
Forest-%	✓	✓	✓
Agriculture-%	✓	✓	✓
Most petro locations	✓	✓	✓
Most diamond locations	✓	✓	✓
Most elevation	✓	✓	✓
Most forested	✓	✓	✓
Most agriculture	✓	✓	✓
Most attacks in current area	✓	✓	✓
Greatest area	✓	✓	✓
Aid per location, t-1	✓	✓	✓
Total funding, t-1	✓	✓	✓
DV lag1	✓	✓	✓
DV lag2	✓	✓	✓

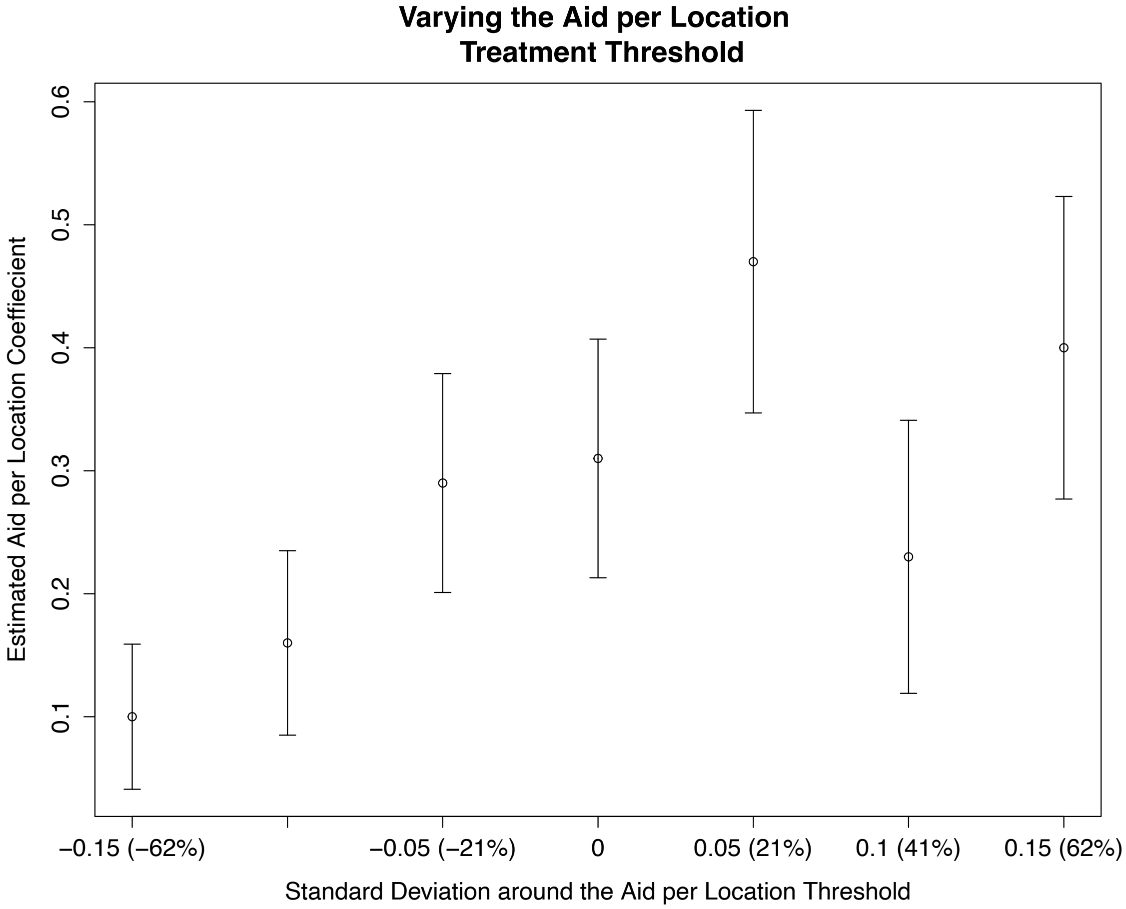
Year dummies (1989-2008)	✔	✔	✔
Obs.	2372	2372	2372
Treated obs.	370	370	370
Matched obs.	2362	2359	2359
Matched unweighted	4603	4526	4526
Caliper (SDs)	0.1	0.1	0.1
Obs. dropped by caliper	10	13	13

The Average Treatment Effect of 0.31 represents a 52% increase in military fatalities (log) in treated observations compared to control observations (mean=0.6 in unmatched sample). In actual numbers of fatalities, that represents a 138% increase, or an additional 4.1 fatalities compared to the 2.98 fatalities if there is no, or low, funding concentration.[2]

The main result is robust to many alternative measures of the most important independent, dependent, and control variables. Starting with the independent variable we checked whether our findings were robust to shifting the threshold for when an observation is considered to be subject to treatment. The different thresholds vary between -0.15 and 0.15 standard deviations around the mean of Funding per location (see Table 1). Figure 5 illustrates that when analyzing funding concentration's effect on military fatalities based on alternative thresholds, the results remain but the effects and statistical significances decrease at 38% ($257,336) of Funding per location's main cut-off point ($676,274).

Figure 2: Displaying the impact, and confidence interval, of Funding per location on military deaths, as a result of varying the treatment threshold

[2] We transformed the ATE from logarithms to numbers by first adding the ATE to the baseline (average DV for control observations, IV=0). We raised 10 to this sum (10^0.91) and from that figure subtracted 1 to arrive at the sum in actual fatalities (7.1). By subtracting the baseline in actual fatalities (7.7-2.98) we arrive at the ATE in fatalities rather than log of fatalities. It is then possible to calculate the percentage increase in actual fatalities from the baseline of fatality numbers (4.1/2.98=1.38).



We consider that the Funding per location measurement, of the currently available options, best represents the theoretical concept of funding concentration. We did however also check the results when using the alternative measures, Aid per square kilometre and Total funding. The directions of the effects remain but the statistical significance levels drop to the 90% level. The theoretical impact of Total funding, irrespective of its concentration, could be interesting to develop in its own right. For now, the results are strongest for the Funding per location measurement.

We also check whether the results are robust to an alternative specification of one of the most important control variables, Battleground control. When replacing this measure the results are essentially the same. Although we already control for whether an administrative division contains a capital we also checked whether the results hold when all capital regions are completely excluded from the dataset and found that the effect increased while remaining at the 99% level.

It is possible that how warring parties expect that aid will be implemented has changed over the years. If lessons were learned – for instance since the early 90's humanitarian operations in DRC (Polman, 2010), and food aid operations in Somalia and elsewhere (Addison et al., 2002, p. 383; Maren, 2009; Natsios, 1996) – we would then expect the impact of highly concentrated aid to be ameliorated later in the dataset. We selected 1997 as the starting point since by that time policy makers should have become aware of the problems resulting from when donors created pockets of highly valuable aid in the early 90's. Andrew Natsios (1996) for instance suggested that donors flood markets with food aid to diffuse its value to warring parties. We find that the direction of the effect remains, albeit lower (0.22). The statistical significance dropped to the 90% level. This might suggest that aid implementation has improved over time, but could also result for technical reasons. We for instance know that the quality of available aid data has increased over time. Although there are related studies supporting the idea that funding concentration can fuel violence, this field of inquiry is still in its infancy and will benefit greatly when more geocoded aid data becomes available from more sources, for additional time periods and regions.

A Preparing the Data sets

We make use of novel geo-referenced *aid* (Author, 2011) and *battleground control* data, paired with *UCDP's Geo-referenced Events (UCDP-GED)* dataset (Sundberg & Melander, 2013). The original datasets are of events format and we have aggregated this information into annual data. While doing this we also aggregate into administrative divisions. There are hence many involved steps before the data can be analysed. The purpose of this appendix is therefore to communicate exactly which countries and years the data covers; which types of administrative regions that are included and why we rely on these political boundaries rather than gridded cells; how the original *aid* events data was collected and aggregated; how the *battleground control* data relates to the original UCDP-GED events and how it was coded and aggregated.

A.1 Time Period and Spatial Boundaries, Case Selection[3]

The dataset makes it possible to analyze how aid, battleground control, as well as different geographic factors, affect fatalities in administrative divisions. The biggest caveat is that the dataset only covers areas that are *violently contested*, meaning that at least one person died in the area a given year. It is therefore not advisable to test broader hypotheses concerning funding that is committed to uncontested areas. The focus on already contested areas also makes it impossible to test hypotheses concerning the *onset* of violence.

To enter the dataset a country in Sub-Saharan Africa needs to have experienced at least one year of intra-state conflict since the start of 1989. To experience intra-state conflict means that there have been 25 annual deaths, or more, in violence between at least one organized group and the government. The thresholds and year ranges are based on standards adopted by the UCDP (Harbom et al., 2007).

Once a country has entered the dataset, the years where *organized non-state* groups battle each other are also coded. Years when there is only violence between *unorganized communal* groups are not coded. A country leaves the dataset as soon as there is a year without intra-state conflict or non-state violence.[4] After spells of inactivity a country can always enter the dataset again.

A country in which a conflict starts after 2007 is not included. For the cases that are included in the dataset, 2008 is the last year that is coded.

[3] Parts of this section has already been made available online in Author (2012).

[4] The main interest here is to include organized violence that occurs in the shadow of state-based conflict. To that end we would also include some state versus state violence, as long as it results from the presence of an intra-state conflict. An example of that type of violence is the combat between the government of Ivory Coast and France in 2004 during the Ivorian intra-state conflict, which would have been included had there been high enough level of fatalities.

The geographic extent of what is here called Sub-Saharan Africa includes a couple of cases that are sometimes not considered a part of Sub-Saharan Africa, and lacks some candidates. Table A1 for example includes Sudan but excludes Mali. The list details which countries and year ranges that are covered.

Country	State-based conflict years	Additional Organized non-state conflict years	Comment
Angola	1989–1995, 1998–2002	--	
Burundi	1991–92, 1994– 2006, 2008	1991–92, 1994–2006, 2008	
Central African Republic	2001–2002, 2006, (2009–2010	--	
Chad	1989–1994, 1997–2002, 2005–2010	--	
Comoros Islands	1989, 1997	1998	
Congo, Democratic Republic	1996–2001, 2006–2008	2002–2004	
Congo, Republic of	1993–1994, 1997–1999, 2002	--	
Djibouti	1991–1994, 1999	--	Versus Eritrea 2008 not coded since it is an interstate conflict

Eritrea	1997, 1999, 2003	--	Djibouti versus Eritrea 2008, and versus Ethiopia (1998– 2000) intermittent year 2001 not included since those are interstate conflict years
Ethiopia	1989–1996, 1998–2008	--	
Guinea Bissau	1998–1999	--	
Guinea (Conacry)	2000–2001	--	
Ivory Coast	2002–2004	2005	
Lesotho	1998	--	
Liberia	1989–1990, 2000–2003	1991–1992, 1994–1996	
Mozambique	1989–1992	--	
Nigeria	2004 (2009)	2003–2004 (2008)	Conflict year 1996 (Cameroon vs Nigeria) is not coded
Rwanda	1990–1994, 1997–2002, (2009–2010)	--	
Sierra Leone	1991–2000	--	

Somalia	1989–1996, 2001–2002, 2006–2010	1997–2000, 2003–2005	1998–2000 are missing since non-state prior to 2002 weren't available when we started coding
Sudan	All years	--	Note that South Sudan was considered a part of Sudan by the time the area was coded
Uganda	1989–1992, 1994–(2010)	1995–1997, 2003, 2004	

Table A1: Countries South of the Sahara with Active Conflicts 1989–2008

A.2 Why Administrative Regions over Grid Structure?

There are a number of ways to structure sub-national datasets when investigating conflict outcomes ranging from using existing political boundaries to creating entirely exogenous spatial divisions in the shape of quadratic grid cells. When they released a standardized grid structure, the PRIO-GRID, Tollefsen, Strand and Buhaug (2012) reviewed the pros and cons of these different methods.

The main benefit of gridded cells is that they are entirely exogenous to conflict dynamics, whereas administrative regions could actually result from underlying conflicts or power distributions. Gridded cells are by definition the same size across time and space whereas administrative divisions vary between areas and across time (Buhaug & Rød, 2006, p. 322; Tollefsen et al., 2012, p. 365). When using administrative divisions it is for instance necessary to control for the size of the area and related variables.

Administrative regions can at most be analyzed at two or three scales, from first order administrative division to second or third order divisions. In contrast, a gridded structure is easily scalable up and down in continuous steps, thereby making it straightforward to determine the range of scales over which results hold (Tollefsen et al., 2012, p. 365).

Despite the compelling reason for using a grid structure other approaches have been and continue to be used (Buhaug & Lujala, 2005; Østby, Nordås, & Rød, 2009; Weidmann & Ward, 2010). We realize that there are good reasons to create a dataset using the PRIO-GRID structure but for our first analyses we have three reasons for using a administrative division structure. Firstly, the main data that we depend on – that which contains aid and conflict information – is coded in relation to administrative region. By aggregating data within administrative divisions rather than grids we are able to keep more data without introducing assumptions about how to divide administrative region data over cells.

Secondly, we find it more straightforward to relate our theory to the data when using administrative divisions. Our model assumes that a warring party has different strategies for different politically relevant areas. The theory is not restricted to a specific type of politically relevant area, but we find that it makes sense intuitively to make use of a first order administrative division structure. This is an area that is politically relevant and large enough to warrant an independent contest strategy, while being small enough to provide sufficient within country variation for statistical analyses.

Thirdly, by using administrative divisions it is easy to pull out examples from the dataset with actual names. This makes it easier to relate results to case studies and to communicate findings outside of academia.

Administrative divisions in Africa South of the Sahara are subject to change over time. To make sure that the units of analysis are constant over time we use the latest administrative region data from the Database of Global Administrative Areas (GADM, 2013) and let the current pattern of administrative divisions represent all years in the study. This means that aid going to, for example, the Ethiopian province *Eritrea* in 1989, would in the static dataset be counted as going to the Eritrean region *Debub*. The region Debub contains the central point, i.e. the pair of coordinates that represented Eritrea when it was an Ethiopian province.

The decision to use units that are constant over time does not result in significant data preparation problems. It is, however, crucial that users of the data realize that the names of regions only reflect the most recent divisions. The name of funded areas may be different in aid project descriptions compared to the area in which information is aggregated, but the geographic location is approximately the same.[5]

The focus is on first order administrative divisions like provinces. Although it would be possible to achieve greater granularity by focusing on lower order administrative divisions the higher level is selected to minimize the number of observations where there are zero events. Having too many zero observations limits the number of different methods of analysis and limits the number of control variables.

There are a few cases, such as Sudan, Chad and the Democratic Republic of Congo, with administrative divisions that are registered by GADM as first order, but that are much larger, and often fewer, than those of most other countries. For these deviating cases the areas recorded by GADM as second order administrative divisions are used instead as they are more comparable to the sub-divisions of the other countries.

A.3 Geo-referencing Foreign Aid[6]

Having established the scope and structure of the dataset we now turn to its actual content starting with geo-referencing aid events. Prior to our (Author 2011; Author, 2011) geo-referencing work, very little subnational aid data have been available. Some donors have geo-referenced their projects on a limited scale, including some country offices of the United Nations Development Program (UNDP), which have completed city-level coding in Kenya and district-level coding in Nepal.[7]

The Mapping for Results partnership between AidData and the World Bank provides the most widespread donor-generated database – by 2011 active projects had been mapped in more than half of the World Bank partner countries – but even this data set has been

[5] That is, within the margin of error that are implicit in the precision coding. The pair of coordinates that represents the center (or rather the so-called centroid) of a province named in an aid project can be shifted if the province gets slightly different geographic extension in the static version of the administrative division data. The precision code 4 does imply that the pair of coordinates that best represent the area could be anywhere within a province sized area.
[6] This section contains text that has been published in Author (2011) and Author (2011). The former should be cited when using the geo-referenced aid data and the latter when using the coding methodology.
[7] See http://www.undp.org.np/index.php for UNDP's active project database in Nepal.

restricted to active projects.[8] The Aid Locations during Civil Wars South of the Sahara dataset (Author, 2011) represents the only historic sub-national dataset currently in existence.

A.3.1 Subnational Geo-referenced Foreign Aid Events

The geo-referenced aid projects are drawn from events in the AidData core dataset. AidData core contains funding commitments since 1945 from most multilateral and bilateral donors.[9] Because of our interest in the aid-conflict relationship, we prioritized geo-referencing aid projects since 1989 that are committed to African countries in which there are ongoing armed conflicts.

The system of geo-referencing used here is based on the Uppsala Conflict Data Program's *Geo-referenced Events Dataset* (Sundberg et al., 2010) and has been adapted to the specific coding decisions that need to be made when geo-referencing aid projects (Strandow et al., 2011).[10] The system distinguishes between pairs of coordinates on four main levels of precision, ranging from point locations, through two administrative divisions, to the country level. In addition to the four main precision categories there are four additional codes to further separate different levels of certainty in the coding. The criteria for the precision codes are as follows:

> 1–2: Used when a location lies within (1) or near (2) a specific populated place or object.
> 3: Used for a district or municipality.
> 4–5: Used for a specific province (4) or a greater region (5).
> 6: Used when a project is national in scope.
> 7: Used when no location is given or location is unclear.
> 8: Used when aid flows directly to a government entity.

[8] See maps.worldbank.org for the World Bank's active project database in Sub-Saharan Africa, Latin America, and selected Asian countries.

[9] AidData primarily contains foreign aid commitments (Tierney et al., 2011). When using commitments, the question of whether the committed aid actually arrives in country is always open. Unfortunately, few options exist to remedy this concern. Disbursement data is not extensive, and according to the CRS (Creditor Reporting System), it should not be used before 2002, given how inconsistent the data are. According to the OECD: "the analysis on CRS disbursements is not recommended for flows before 2002, because the annual coverage is below 60%...."

[10] For most recipient countries experienced coders considered each project only one time in this iteration of the dataset. Cases with a great number of aid projects or cases that proved particularly difficult to code, like Ethiopia, Somalia and Uganda, were resolved by collaboration between at least two coders.

Using precision codes makes it possible for users of the dataset to select subsets that contain different levels of precision. Sources vary greatly in how precisely they record geographic information; sometimes the exact location is named, in other instances the general area is reported, while sometimes much of the country is the intended beneficiary (such as for a program to combat AIDS/HIV for much of the population).

Foreign aid projects granted to a national government entity, for example, may intend to reach beneficiaries unevenly distributed throughout the country. Because we cannot assume that the entire country receives aid, the precision code of 8 signifies that the money flows through the government entity, but is unclear afterward. Users of the data could keep such projects in their analysis, or remove them if desired.

The coders tasked with geo-referencing aid projects have relied on three main columns in the AidData portal (Tierney et al., 2011) for location information: The project title, project description and short description. The benefit of this approach is that it has been possible to code a vast amount of projects. The downside is that many projects lack descriptions. This – and the fact that we can confirm commitments rather than disbursements – means that the current version of the data is best suited to measuring *expected* aid flows rather than purely material effects.

The core of geo-coding aid projects is to find the name of the location that aid is committed to and then to look up the pair of coordinates that best represents that location in a gazetteer like *Geonames* or the *GNS*. Each row in the AidData portal represents a project, i.e. a funding commitment from one donor to a main economic sector. A project may cover several locations. Coders therefore record the coordinates of all locations in separate columns.

After geo-referencing is completed the dataset is transformed so that a multi-location project is divided into several rows instead of relying on different columns. Each row then becomes a funding commitment to one location. In order that data users can avoid aggregating the same funding over and over again coders note how many locations a project covered. That makes it possible to calculate the average funding committed to each project, or to estimate the expected concentration of aid values. For the complete coding rules see Author (2011).

Figure A1 shows the geographic coordinates we had coded by 2011 for all the countries with conflicts throughout Africa using all precision codes except 7, the unclear cases.

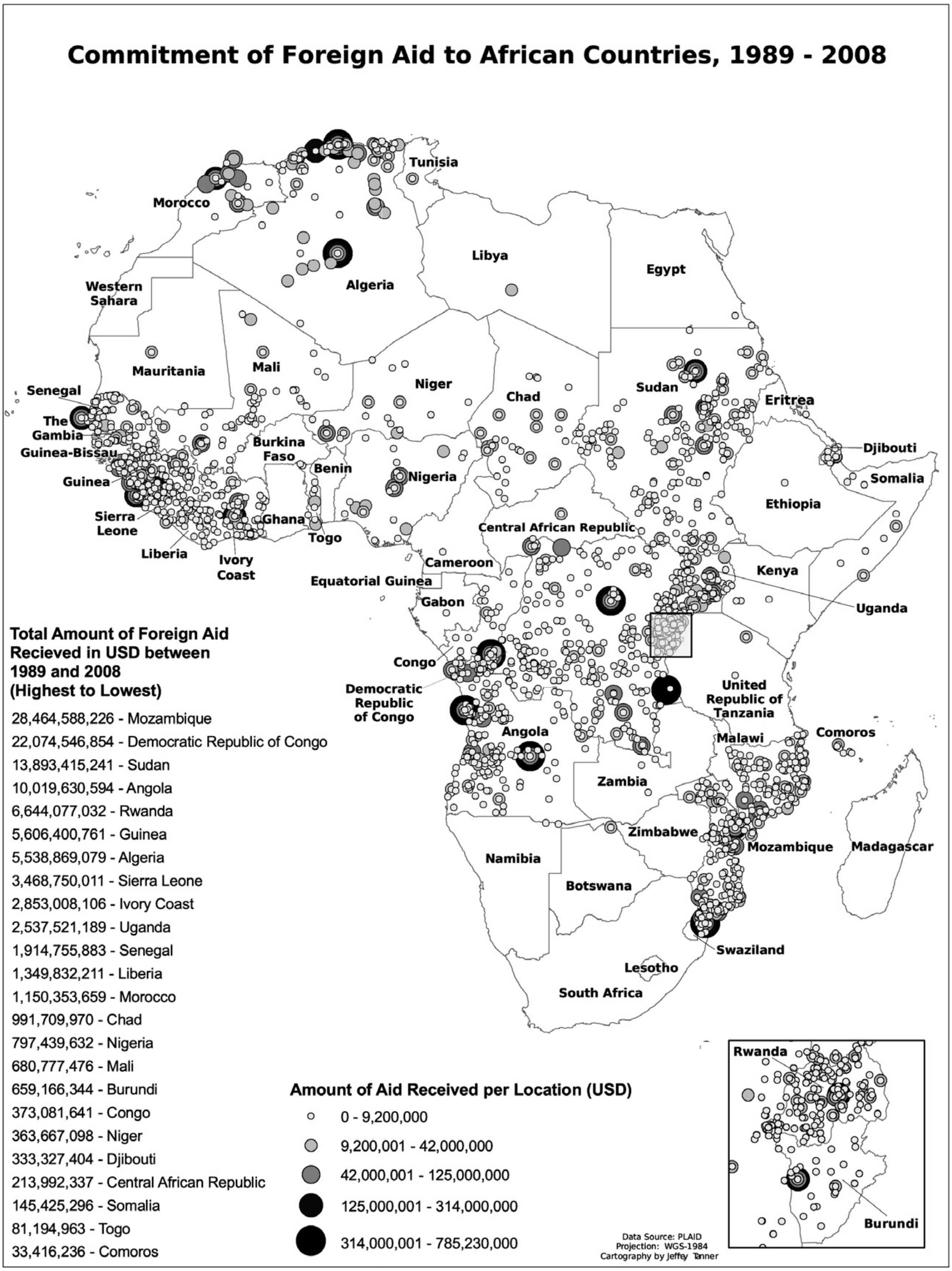


Figure A1. This map contains all aid projects that we had geo-referenced (assigned geographic coordinates) based on project descriptions by 2011. Each dot on the map represents a discrete aid project and is scaled by the amount of aid it represents as depicted in the legend. From (Author 2011).

A.3.2 From Aid Events to the First Order Administrative Division

We have so far clarified for which countries that events are coded; the geographic level of analysis; and the basics of how aid events are geo-referenced. We now specify how the events data are aggregated to the *year* and *area* format. It is not until later when we introduce the battle data that we add the additional *warring party* dimension. For now the year and area dimensions suffice for defining the unit of analysis.

The aid events data, in Excel format (Author, 2011), is opened in a Geographic Information Systems program, ArcGIS. The dataset contains geographic coordinates for each row so that the events can be plotted as point data over the static administrative division map. There is now one layer with point data and one with polygon data. The aid points that have an accurate enough precision score are selected and joined spatially with the administrative division file. This results in excess of 19,000 rows that has inherited the properties of both points (precision scores 1, 2, and 8) and polygons (3, 4, and 5).[11]

The administrative division data contains information about the names and ID's of the administrative divisions. The aid events comes with information about the year that funding was committed; the amount of funding; the sector that was funded; and the number of locations that was funded. It is this information that needs to be aggregated.

Before aggregation we calculate the event version of funding concentration by dividing the constant 2000 US dollar value (Author 2012) by the number of locations that a project was committed to. This is done so that aid values are not double-counted when summed for different areas. Do note that by avoiding this summation problem the assumption is introduced that a projects different locations receives the same share of the funding.

Aggregating the events is done by what is called the *dissolve* command in ArcGIS.[12] All relevant variables are summed over the two identifying variables *year* and *area*. The events

[11] Precision score 8 is considered high enough precision to be included. We include precision score 5 projects since the first order administrative division that lies in the centre of a greater (precision 5) area is usually a good approximation of where projects were located. It may not always be advisable to include precision score 5 projects into administrative divisions, it depends on what theory is being tested.
[12] From this section and onwards we use some terms that may seem convoluted, like *dissolve*, *spatial join* and *union*. We go with these terms as they are used in ArcGIS to perform the described operations.

format is thereby dissolved into a *year-area* structure. The variable that measured funding per location in the events format has now become a summary of all aid value in an area, *total funding*.

In order to arrive at the final measure of *funding concentration* total funding is divided by the total numbers of locations that received aid in an area. Fewer locations gives greater expected funding concentration. It is also at this point that *funding per area* is calculated by dividing total funding by an area's size in square kilometers. All of these funding measures are then transformed into dichotomous versions.

A.4 Coding Points of Control and Attack[13]

It is now time to clarify how we determine battleground control. We first describe the events dataset that contains information on which locations that warring parties attack and control. Then, in section 4.5, we relate how to go from these events to the yearly administrative division format. It is at that point that the warring party dimension is added to the year and area structure.

There are four main steps of coding and re-coding to arrive at the events dataset. First, the Uppsala Conflict Data Program geo-references battles and collects information on fatalities and other aspects of the clashes (Sundberg & Melander, 2013). The UCDP-GED is of a so-called undirected dyadic format. A dyad is a pairwise interaction between two warring parties, or one warring party and its civilian victims. That a dyad is *undirected* means that a battle between two warring parties is only recorded once. Any information that differs between parties is recorded in separate columns. For instance fatalities that party *A* suffers is listed in a column for *A* deaths and for the *B*-side there is a column for *B*-deaths. A *directed* dyad format would on the other hand have two rows for one battle. The fatalities for both sides would then be collected in the same column but in two rows.

For internal UCDP use the circumstances surrounding a clash are saved in a comment column. This is where UCDP's work ends and where the second step begins. Coders go through the text in these comments and disentangle the narrative of violent events. They read

[13] Parts of this section has already been shared publicly (Author, 2013).

the battle descriptions and determine which actor that initiated an engagement. Following each violent incident the territory or the object in dispute can be defended or change hands, and in the end either party may control it. The coders are hence able to determine so called *points of control*. Information about control is in general so sparse that it is not viable to implement rules for how long time a party must possess a location after battle in order to be coded as in control. As long as it is not obvious that a party takes control and immediately moves out and gives up the location, then control is coded 1 for the controlling actor.

The third step of preparing the dataset is to transform it from an undirected to a directed dyad format. This is done because the final dataset focuses on one warring party at a time, and not on separate dyadic interactions.[14] The transformation is done by a Java-script.[15] All original columns are saved after this transformation, which means that it can easily be treated as an undirected dyadic dataset by toggling the *directed* mode on and off.

When using the dataset in the directed dyad mode there is one column for attacks (*A* attack) and one for control (*A* control). When coded 1 the actor that is labelled party *A* asserted control (*A* control=1) as a result of the clash. If the variable is coded 0 then the coder(s) determined that the party did not assert control. In the situation where both parties have partial control of a location then *A* control is coded 8. If coded 9 then there was not enough information to code whether the actor achieved control or not. The same codes are used with analogous meanings when coding who initiated the attack.

The rules for coding which side that asserted control are designed to deal with state-based and non-state events. For one-sided violence the attacker and recipient are per definition *A* attack=1 and *B* attack=0. For information on types of violence see for instance Harbom et al. (2007). For the full rules of coding straightforward as well as difficult cases see Author (2012).

The final preparation of the events dataset before aggregation is to make it possible to summarize the control (and attack) columns over time and space by dealing with the more

[14] For future uses of the events dataset it is also much easier to complete advanced transformations of the data in ArcGIS with a directed dyad format. For instance when drawing areas of control for separate warring parties (Strandow et al., 2013).
[15] The Java script was coded by Suzuki XXXX.

ambiguous categories 8 and 9. In essence the prepared *A* control column is coded 1 if *A* control is 1 *and not* 0, *or* 8, *or* 9.

A.5 Establishing Battleground Control in Administrative Divisions

The focus of this appendix has so far been to explain how aid events were aggregated to a yearly area format and to explain how the battle control events dataset was structured. The events dataset results from recoding the UCDP-GED, which means that all the information concerning fatalities (and thereby violence intensity) that is contained in that original dataset, is kept in the battle control events dataset. Due to the similarity between the control and fatalities columns, when we detail how the events dataset is aggregated into yearly administrative divisions we indirectly relate how event fatalities are aggregated into violence intensity. In the next section (4.6), where we explain how the aid and battle data is combined, we further discuss how violence intensity is aggregated.

The purpose of this section is to establish the procedure for measuring a warring party's *battleground control*. To recapitulate, battleground control is the extent to which a warring party has been successful in winning battles by defending or conquering territory within a sub-national administrative division, such as a province.

The concept *contested areas* is key to understanding what battleground control is and what it is not. We are not measuring territorial control resulting from conquest that was unopposed by other warring parties. For instance a big part of the initial conquests by the rebels in Ivory Coast in 2002 and the advances by the rebel group MODEL in Liberia in 2003 do not count towards their battleground control, as there were few clashes with government troops at the start of the conflicts. We only gauge the control that was revealed through violent contacts significant enough that at least one person was reportedly killed.

A.5.1 Aggregating Battleground Control

The battle control events data that is used to generate the battleground control dataset has a directed dyad structure. Since the end goal is to have a party focus we transform these directed dyads into a *year-area-party* structure. This unit of analysis focuses on one warring party at a time and its interaction with all other warring parties in an area. All parties in such a conflict cluster are investigated in turn. To summarize, in the resulting data structure each row represents a *warring party* in a particular *area,* for a particular *year*, against all opponents in that area.

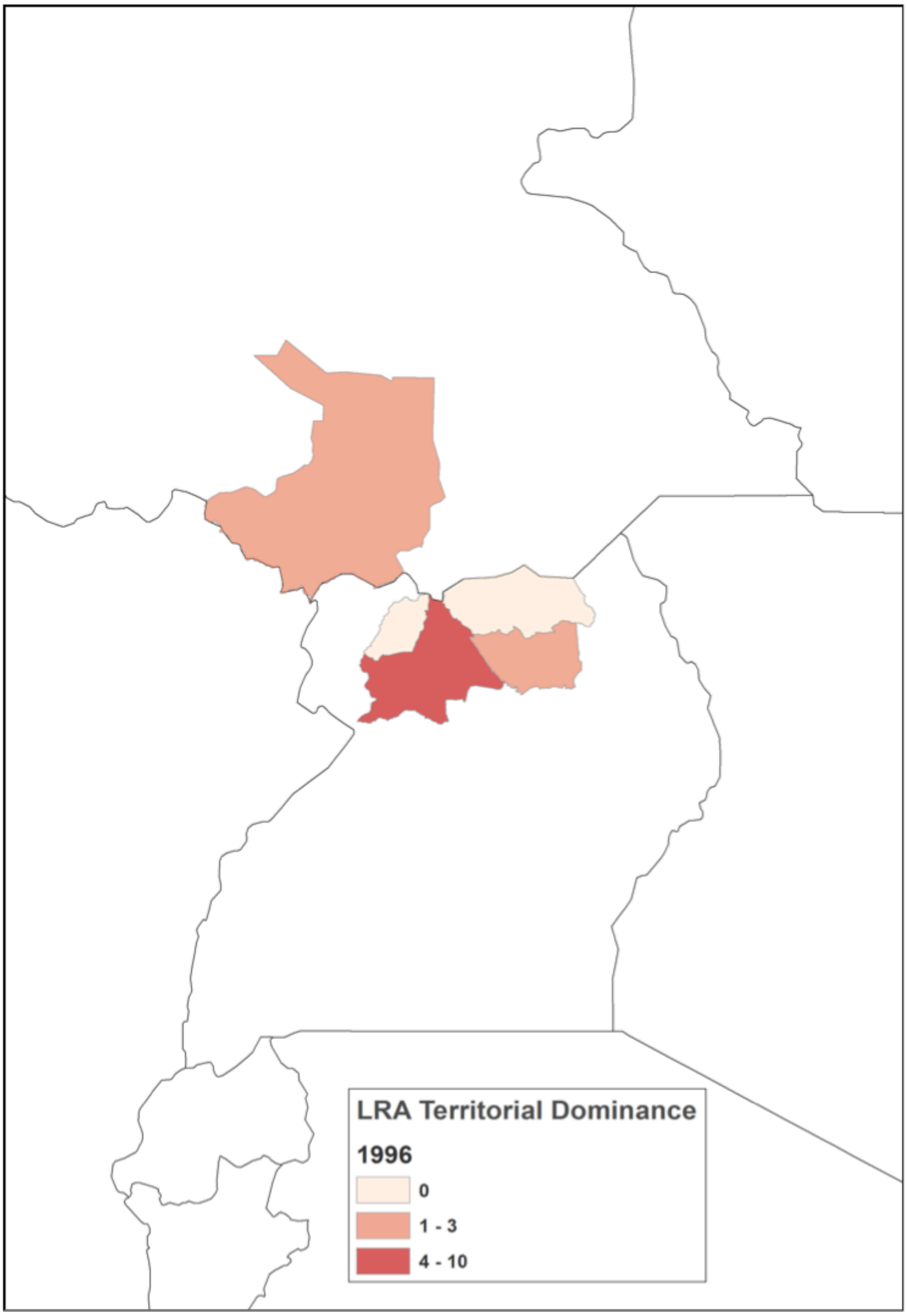


Figure A2: Battleground control by LRA in Uganda and South Sudan. Uganda is in the center of the map and South Sudan is to the north-west. The deeper the color the greater the battleground control, areas that are not coloured are not violently contested by LRA that year

Figure A2 displays battleground control for the Lord's Resistance Army (LRA) in 1996. That year LRA reportedly took control over locations ten times in the Gulu Province in Uganda after attacks against civilians and battles with the government. Following clashes with civilians and the Sudan People's Liberation Movement (SPLM/A) the LRA also took control over one location in Central Equatoria in what is now South Sudan.

A.5.2 How Battleground Control is Established

This is a brief description of how battle events are aggregated into battleground control. An Excel file containing directed dyad events data is opened in ArcGIS or similar program. Events with a spatial precision of 4 or lower are exported to a new file and are plotted over the map of administrative regions. This is a copy of the original static administrative division map to which aid has not been joined.

First order administrative divisions have precision score 4, and the precision scores greater than 4 generally refer to clashes whose locations are more imprecise and that can only be related to bigger geographic areas (or that occur at sea). By excluding more imprecise events only battles that clearly occur within the area of analysis are recorded.

Before joining the events to the map the battles are related to estimations of populations size at the closest populated places, if there is any within about 10 kilometers from the site of battle. This data comes from CIESIN's *GRUMP settlement points* (CIESIN, 2004) and is used as a measure of how signifiant a battleground is.

As with the *aid* events, the *battle* events are spatially joined to the administrative divisions map and the resulting file is then dissolved in order to establish the yearly area data. Unlike the aid data the battle events are also dissolved over the warring party's identification number (as assigned by the UCDP-GED). Statistics are generated for many variables including control points. The sum of points that a warring party *A* has asserted control over in an area for a particular year becomes that party's battleground control in absolute terms. The sum of points that all opposing parties have controlled during the year are aggregated as side *B*'s absolute battleground control. Side *B* no longer refers to a single opponent but to all opponents in an area. It is now possible to calculate *greater battleground control* in both ratio and dichotomous formats.

A.6 Getting it All together

We have so far explained how aid and control events were aggregated into *year-area*, or *year-area-warring party* format. We will now discuss the dependent variables and the temporal lag structure. We thereafter relate how data of raster formats were added to the final dataset and lastly we describe how all datasets were combined and how a spatial lag structure was set up.

A.6.1 Violence Intensity and Temporal Lags

We have already stated that the procedure for establishing battleground dominance in administrative divisions was also used to produce aggregates of fatality figures. This was done at the point that the *dissolve* command was used to set up the area-year-warring party structure. It is at this point possible to specify a number of ways that events are summarized into aggreagates. Means, sums, counts, and other statistics can be used. The *sum* statistic was used for aggregating civilian deaths, unknown deaths, the fatalities that both sides *A* and opponents *B* suffered, and total fatalities.

It was also at this point that other variables related to the warring parties were aggregated. A variable that records whether a party can be termed challenger or incumbent was aggregated using the *mean* statistic. All parties have only one role at a time in a particular area, either incumbent or challenger, but if they would be both incumbent and challenger at the same time in a particularly complex situation that would have been captured by using the mean statistic.

To ensure that the dependent variables occur in time after the independent treatments, and to make it possible to control for earlier values of the variables, we added temporal lags and leads. In order to easily verify that the lag structure was correct this was done in both ArcGIS and Excel in several steps. The basic procedure is to combine rows from year t+1, or t-1, with row t. This was done by creating three IDs for each row that combined (i.e. concatenated) a row's year, country, administrative division, and warring party IDs. The three ID's differed only in that one contained year t, the other t+1, and t-1.

The end result is that it is possible to, for instance, add the total deaths in battles between LURD (Liberians United for Reconciliation and Democracy) against the incumbent in Montserrado (the first order administrative division that contains the capital) in 2004 to the independent variables in 2003 (for the same party in the same area).[16]

A.6.2 Adding Raster Data to the Administrative Regions

This section is not required for most readers except those that want to get a sense of how some of the control variables were generated. Raster data represents spatial information in a matrix of cells. Polygon data in gridded format, like PRIO-GRID, look like raster data but is different since the former usually connects a number of columns of data to a map whereas for a raster, the map is essentially the data. For instance population size in raster format consists of cells of different colors where each color corresponds to a value. In polygon format each cell could display a number of different variables besides population size, for instance population density.

There are quite a few variables that are of interest to conflict researchers that are only readily available in raster format and that need to be transformed into polygons. Here rasters have been transformed either by converting a raster to polygons or by creating zonal statistics and adding those statistics to existing polygon data. Selecting a transformation method depends on the type of data and what information that needs to be transferred from the raster to polygons. We start by describing the convert raster-to-polygon method and the follow up procedures used for generating yearly population data. We then describe the zonal statistics method used for connecting mountainous terrain and land use rasters to polygons.

The *convert to polygon* method was used to convert population rasters (CIESIN, 2000) to polygons.[17] Neighbouring cells with the same population size are thereby converted into one polygon. It is then possible to multiply the number of cells by the cell value (population size)

[16] That combines the independent variables of time t with the dependent variable at t+1. Note that this is done, quite counterintuitively, by merging the t-1 ID in 2004 (which contains the year 2003) to 2003's t ID (which again contains 2003). It seems convoluted but it makes the actual merging easy.
[17] This operation was performed by Miguel A. Pavon, Adjunct Professor, University of Texas at Austin.

to calculate the total population in a polygon. These polygons were then put in union with a copy of the administrative division dataset and the parts of the population polygons that fell within an administrative division was counted towards that area's population. This was done in a way that ensured that no population was double counted or under counted. The procedure was repeated through the three sets of population estimates released by CIESIN, covering 1990, 1995, and 2000. These three datasets were then spatially joined to the administrative division dataset so that data from the three time periods were added as three columns.

Since the final dataset is yearly this population dataset was then joined to a yearly version of the administrative divisions dataset. To make this more manageable the population polygon data was transformed into points. This transformation to points introduces no errors in the location of population sizes. The points were then spatially joined one-to-many with the target feature, the yearly administrative division dataset. This procedure ensures that the population records are copied to all yearly observations of the administrative division data.

The population data is still not ready for temporal analysis. Recall that the temporal information is included as separate columns. The temporal information needs to be turned into one column that varies over three periods of time (1990, -95, and -00). We therefore created a new population size column and copied the records from 1990 to all years ranging from 1989 to 1994. We did the same with the 1995 records for the 1995 to 1999 years. The years from 2000 and onwards were dealt with in the same way. There was however some missing values in the 2000 records that were filled with information from 1995. Following these procedures the population data was prepared for temporal analysis.

In order to calculate population density the yearly population dataset was projected to Eckert IV so that the administrative divisions' areas could be calculated in square kilometres. This makes it possible to calculate a standard measure of population per square kilometre. Before merging with other datasets the map projection was converted back to the standard WGS 1984 format.[18]

[18] There were some random missing variables throughout the population data so the final variables that are used in analyses here had missing information replaced by area summaries of the population point dataset (CIESIN, 2004) used for determining the significance of battle locations.

Turning now to the second method. Landcover (European Space Agency, 2012) and mountaneous terrain (UNEP-WCMC, 2002) values are ordinal scale variables in their raw format. Ordinal scale values have a ranked order between values but ratios of values cannot be meaningfully compared. The terrain variables can be recoded into ratio format but they first need to be transformed into polygons. The raster-to-polygon procedure worked well with population data due to its ratio scale format but the terrain measurements need the second procedure, *zonal statistics as table*, which can summarize statistics from ordinal varibles. The table of statistics can then be associated with a polygon dataset.

To retrieve statistics from a raster into a table we again relate the raster to the administrative division dataset. With the exception that we use a version of the administrative division dataset that only includes areas where there have been some form of violence (the aforementioned area-year-warring party structure).[19]

The summary statistic that captures the central tendency, *majority*, was used. This statistic measures which cell value that is most common within an administrative division. Using the join field command the summary statistics is joined back to the full, yearly, shapefile.

A.6.3 Combining all Datasets and Creating Spatial Lags

Similar to how the temporal lag structure was created it is also possible to prepare and merge variables from different datasets and to create spatial lags. We start by describing how the aid and terrain variables were added to the conflict dataset, and end by explaining how the spatial lags were set up.

In all datasets one new ID column was created that combines the year and the country and region ID's. The aid variables were then merged to the conflict dataset. Note that all aid committed to an area will be added to each warring party observation in that area. It is therefore not possible to aggregate aid to the country level based on this dataset, as that would inflate the total amount of aid committed to a country. The different terrain and rainfall variables were merged to the conflict dataset in the same way.

[19] Note that this procedure produces statistics for unique polygons so that the year dimension is lost in translation.

Setting up the spatial lags was done in two ways. The first method finds the maximum value of the variable over both space and time. It compares the current area only to neighbouring areas. It is therefore best for static data of phenomena that are most likely to spread if close to the current area. The second finds the maximum over the entire country but for each year and is therefore good for yearly variables of phenomena that may spread over greater areas than just the closest neighbourhood.

The spatially lagged neighbouring terrain variables that do not change over time are *most mountainous terrain*, *most forested terrain*, and *most agriculture*. These variables were calculated by doing a spatial join between current areas and neighbouring areas and by finding the maximum values of those neighbours. These spatial lags are dichotomous and are coded 1 if current area has as great, or greater, figures than the neighborhood maximum.[20]

The second method illuminates general diffusion with the variables *A over peer attacks*, *most diamonds*, *most petro* and *greatest area*. To calculate these lags the first step was to concatenate an ID that combines year, country, and warring party IDs. Leaving out the area ID means that the resulting dataset focuses on all areas in a country in which a warring party operates. It is now possible to dissolve a copy of the dataset based on this *year-country-warring party* ID and generate mean values of the variables of interest. These mean values can then be merged back to the original dataset. Now the general diffusion can be determined based on the difference between values in the current area and the country means.

Both of these methods will only reflect spatial lags, or general diffusion, between contested areas since there are no uncontested areas in the dataset. There are definitely reasons to revisit these methods in the future as, for instance, if rebels have safe havens (due to heavy forestation or elevation) in neighbouring uncontested areas, that could affect strategy and violence in currently contested areas.

Having added spatial lags the dataset has all variables necessary for analysis, independent, dependent, and control variables, as well as temporal lags. Recall that this final dataset has a year-area-warring party structure and that the B-side consists of all opponents that party a

[20] *As great* or greater is a criterion since the current area is included in the neighbourhood.

faces in a province. *All opponents* include civilians if the warring party has conducted one-sided killings. However, since the criteria for including an area in the dataset is that it has been violently contested, no areas where there are only civilian casualties are included. There has to be at least some level of engagement between military forces for an area to be included.

Appendix B: Control Variables and Covariate Sets

Although matching does not resolve issues with unobserved covariates, in principle matching on an extensive list of covariates could account for much of the influence of alternative factors. What we can do is to be explicit about what controls we are including and what information we are likely to leave out. In this section we specify several sets of controls ranging from those that are theoretically motivated, to those that are necessitated by the research design.

B.1 Conflict Dynamics Controls

There are three types of variables that are crucial to control for: the warring parties' existing control; the warring parties' types; and the number of *B*-side opponents. Another variable that may impact both the independent and the dependent variables is the number of attacks that a warring party is responsible for. If, for instance, a terrorist group initiated many attacks in an area, that could impact donors will to engage in development operations. Conflict dynamics are hence crucial to control for.

Name	**Description**
Greater Battleground Control	*A* Preponderance in Control over Population. Coded 1 if *A* had a difference in population affected by control > 73580 (twice the average difference
Greater Battleground Control, Alternative	For robustness. *A* more Control Counts. Coded 1 if *A* asserted control over more territory than *B* during current year and area
A is Challenger	Whether *A* is a challenger
Multiple Opponents	Coded 1 if multiple opponents in area
Attacks by A	Sum of all points attacked by *a* in administrative region
A over Peer Attacks	Dichotomous. Coded 1 if current area has as great, or greater, number of attacks by *a* than all other areas within

	the country that actor a operates in
Population near Violence	Mean size of populations at battle locations

Table B1: Attacks, control, and spatial diffusion of attacks

Depending on which treatment is analysed the number of control points and the population size near battles are sometimes included in the covariate sets. We now specify the most important type of covariate, *greater battleground control*, as well as motivate why the type of actor (*A is challenger*) is a crucial control variable.

B.1.1 Greater Battleground Control

The main mechanism through which aid affects violence is the way it nudges a warring party to engage in contests over territorial control in contested areas rather than reverting to irregular strategies. The problem is that warring parties' existing control is likely have a great impact on future propensity to engage in contests over territorial control, as well as on aid commitments. To figure out whether aid has a causal effect it is therefore crucial to include a measure of territorial control when matching observations.

Territorial control could be measured in many ways. Here the main focus is on control in violently contested areas. We are unable to include information from when an area was uncontested as there is currently no systematic source of warring parties control in uncontested areas. That removes the option of estimating the exact distribution of territorial control between warring parties. What is possible to measure is a snapshot of a party's territorial control resulting from previous contests, its *battleground control*.

To set up the *greater battleground control* variable we take into account the value of the battleground control by estimating the population that *A* potentially controlled. The alternative measure is based on the count of the number of locations that *A* had asserted control over.

Greater battleground control attempts to capture the situation where *A* had more valuable battleground control than *B.* To capture this situation the main measure is based on multiplying the number of locations that *A* had asserted control over in an area by the average size of the population in towns within a 10 kilometre radius from that years' battles. The same is done with *B*. We then calculate the difference between *A*'s and *B*'s population control.[21] To make sure that the difference in existing battleground control is substantial enough we code the variable as 1 if *A* had a difference in population affected by control that is greater than twice the average difference.[22] Also, for this variable to be coded 1 both actors need to have asserted control over one person or more. This demand means that only observations where *A*'s control clearly influences more people are coded 1.

The benefit of using preponderance in control over population to measure battleground control is that it is likely to reflect the type of dominance over valuable areas that is most important for subsequent strategic decisions.

The *alternative treatment of battleground control* is coded 1 if *A* asserted control over a greater number of locations than *B*, in a particular year and area and 0 if not. This treatment avoids any potential errors associated with coding population sizes but is more vulnerable to errors in coding control close to zero.

We also include a control variable that measures the *average population near battlegrounds*. Greater population near battles will influence fatalities and may also impact donor commitments.

B.1.2 A is a Challenger

There are a number of circumstances that can push a party to engage in contests over territorial control even if it does no appear optimal in the short term. Incumbents have the overall political goal of protecting and retaking territory, and incumbents' forces are generally geared towards conventional warfare. Furthermore, if an incumbent appears to relinquish its territorial sovereignty to a challenger, without putting up a fight, then more challengers could be encouraged to fight the incumbent. Losing a decisive battle against a conquering challenger may therefore be less costly than the potential reputation costs of yielding or turning to an irregular strategy.

---

[21] The average difference is 36 790. The sizeable systematic difference between *A* and *B* is partly explained by the fact that civilians are never included as side *A*, only as side *B*. This means that there are more *B*-sides than *A*-sides with little or no control.
[22] I.e, > 73 580.

This means that an incumbent should be more likely to expect greater gains from engaging in contests of control compared to a challenger. This is an important control variable since an incumbent is more likely to receive aid.

B.1.3 Multiple Opponents

The presence of multiple opponents is likely to increase violence intensity and may also impact donors' impressions of an area. Coded 1 if there are multiple opponents in the administrative region.

B.1.4 Attacks by *A*

Sum of all points attacked by *A* in administrative region

B.1.5 A over Peer Attacks

This variable records whether party *A* is more active in another contested area than it is in the current area. It is coded 1 if the current area has as great, or greater, number of attacks by *A* than all other areas within the country that actor *A* operates in.

B.2 Resources

Just as foreign aid can influence warring parties' contest strategies so can other valuable resources. We include a number of measures of an area's value ranging from the presence of diamonds to population density. Since absolute values, in percentages or counts, can have diverse effects over different areas we also include relative measures that are coded 1 if the current area is more valuable than surrounding areas.

Name	**Description**
Petro Locations	Number of petro locations within administrative region
Diamond Locations	Number of diamond locations within administrative region
Population Density	Population density
Rainfall	Rainfall in percentages
Agriculture	When there is agriculture land (land used for crops or pastures) indicated at a certain percent, that percent is coded as follows: 150=14%, 140, 180=16%, 110=20%, 120=50%, 20, 30=70%
Most Petro	Dichotomous. Coded 1 if current area has *greater* number of petro locations than all other areas within the country that actor a

	operates in
Most Diamonds	Dichotomous. Coded 1 if current area has *greater* number of diamond locations than all other areas within the country that actor a operates in
Most Agriculture	Dichotomous. Coded 1 if current area has as great, or greater, crops or pastures area percentage than the neighborhood max (since neighboorhood includes current are it has to be "as great or greater" to give similar relation as other resource dummies

Table B2: Resource value control variables

We control for resource value since it affects contest strategy but a wide range of research motivates controlling for resources for additional reasons. It has been found that when there are resources such as gemstones and petroleum within conflict areas conflict duration is greatly increased (Lujala, 2010). We use data on diamond deposits as introduced by Gilmore et al. (2005) and data covering on-shore oil deposits from Lujala et al. (2007).

Availability of resources also reflects an area's income and poverty levels.
Poverty rate, measured as proportion of population below poverty line has been found to increase conflict intensity. Higher poverty supposedly makes recruitment less expensive for challengers. Greater poverty may also drive conflicts due to increasing grievances, and hence support for anti-incumbent activities (Bohara et al., 2006; Do & Iyer, 2010). The resource measure we control for that may take into account some of the population's income is the area's percentage of agriculture land. Rainfall is another measure that is likely to increase the value of land in a way that benefits the general population, while also influencing military operations. More rain makes land manoeuvres more difficult. Rainfall data originally comes from Adler et al. (2003) and is available in administrative region format from Fjelde and von Uexkull (2012).

Population density is used here as a measure of an area's value but different specifications of population has proven important determinants of conflict occurrences. Population size has been shown to increase conflict risks both in terms of populations size at the national level (Hegre & Sambanis, 2006) and in terms of the increased risks of conflicts occurring close to population centres (Raleigh and Hegre, 2009). Common approaches are to measure population as the density per area unit, or as the proportion of a country's total population that lives in a local area compared to the country, or the region in which the capital is situated (Rustad et al., 2011). For population data we use CIESIN GPW Gridded Population of the World (CIESIN, 2000).

B.3 Rough terrain

There are compelling theoretical reasons and anecdotal examples suggesting that rough terrain should influence civil war onset, duration, and intensity. At the country level rough terrain either have no significant impact (Collier, Hoeffler, & Söderbom, 2004; Collier & Hoeffler, 2004), or forests and mountainous terrain have different effects (Fearon & Laitin, 2003; Rouen & Sobek, 2004). In studies that disaggregate analyses to the conflict area there are diverse findings. Mountainous terrain has proven to have no significant influence on conflict duration (Buhaug, Gates, & Lujala, 2009; Lujala, 2010) or a weak positive effect (Buhaug & Lujala, 2005). And forest coverage has been found to not increase conflict onset or duration. (Buhaug et al., 2009; Lujala, 2010; Rustad, Rød, Larsen, & Gleditsch, 2008). In contrast Bohara et al. (2006) investigates violence intensity and finds that violence from both challengers and incumbents increases with rough terrain.

Besides including measures of an area's terrain we also include measures of how rough the terrain is in relation to neighbouring areas. If mountains are used for hiding from opponents it could be important to take into account the potential for hiding outside of the contested area. The viability of receiving aid and the intensity of violence may also be impacted by an area's size and that is hence also controlled for.

Name	**Description**
Mountainous	Real values of minimum elevation in meters
Forested	Percentages of forest cover
Most Mountainous	Dichotomous. Coded 1 if current area has as great, or greater, elevation than the neighborhood max (since neighborhood includes current are it has to be "as great or greater" to give similar relation as other resource dummies
Most Forested	Dichotomous. Coded 1 if current area has as great, or greater, forest percentage than the neighborhood max (since neighborhood includes current are it has to be "as great or greater" to give similar

	relation as other resource dummies
Area Size	Area in square kilometers (based on the area population file)
Greatest Area	Dichotomous. Coded 1 if current area has greater square kilometer area than all other areas within the country that actor a operates in

Table B3: Terrain