Does horizontal education inequality lead to violent conflict?

A GLOBAL ANALYSIS

FHI 360 EDUCATION POLICY AND DATA CENTER

United Nations Children's Fund

Peacebuilding Education and Advocacy Programme

Education Section, Programme Division

Three United Nations Plaza

New York, New York 10017

April 2015

This document was produced through a partnership between UNICEF and FHI 360, as part of UNICEF's Peacebuilding Education and Advocacy Programme (PBEA), Learning for Peace Initiative.

The main authors of this report are Carina Omoeva and Elizabeth Buckner of the FHI 360 Education Policy and Data Center. This study is part of the comprehensive research project exploring the relationship between horizontal education inequality and violent conflict, and the effects of investment into educational equity for peacebuilding, commissioned by UNICEF Peacebuilding, Education and Advocacy Programme and completed by the FHI 360 Education Policy and Data Center research team: Carina Omoeva, Elizabeth Buckner, Charles Gale, and Rachel Hatch. Ania Chaluda developed the back projection module used in the construction of the EIC dataset. Research interns Elyse Sadeghi and Khaled Al-Abbadi provided invaluable assistance in the construction of both EIC and SEIC datasets.

The team is grateful to Bosun Jang and Hiroyuki Hattori of UNICEF PBEA for their continuous guidance and support throughout this project, including helpful feedback in the development of this report. The authors are also grateful to two UNICEF reviewers for their comments and suggestions on earlier drafts of this report.

Contents

Executive Summary 3

Introduction 4

Part I: Global analysis of group-based inequality 5

Dataset construction 5

Descriptive analysis 6

Horizontal Inequality in Education 6

Conflict Onset 9

Regression analysis: Horizontal Inequality and Conflict 11

Covariates 12

Results 13

Interpretation of Results 19

Robustness Checks 19

Part II: Subnational Disparity and Conflict Occurrence in Africa 20

Dataset Construction 21

Descriptive analysis 23

Subnational Inequality 23

Regression analysis: Subnational disparity 25

Results 26

Discussion 28

Recommendations: a research agenda 29

References 31

Appendices 32

Appendix A. Data availability: Global Dataset of Education Inequality and Conflict 32

Appendix B: Sensitivity Checks 34

Subnational Analysis: Regional Variation by Country 37

List of Figures

Figure 1. Distribution of horizontal inequality in education by type 8

Figure 2. Mean group GINI by group type and world region 8

Figure 3. Mean Group GINI by gender and group type 8

Figure 4 . Ethnic inequality in education across time 9
Figure 5. Religious inequality in education across time 9
Figure 6. Conflict onset and incidence, by year 11
Figure 7. Subnational fatalities by year in final subnational dataset 23
Figure 8. Maximum and minimum values in subnational differences from national mean years of schooling, by country 24

List of Tables

Table 1. Number of observations by gender and dimension of inequality 6
Table 2. Summary statistics for Group GINI by group type 7
Table 3. Correlation of inequalities between males and females, by group type 8
Table 4. Geographic coverage and conflict incidence in the UCDP and EIC, by world region 10
Table 5. Country coverage and conflict incidence in UCDP and EIC, by decade 11
Table 6. Descriptive statistics for variables included in regression models 13
Table 7. Model specifications 14
Table 8. Regression estimates: Ethnic and Religious Inequality 16
Table 9. Results of logistic regressions with Subnational inequality as a predictor 18
Table 10. Marginal probability of conflict onset at different levels of horizontal inequality (ethnic or religious) 19
Table 11. Results of robustness checks: alternative specifications of data 19
Table 12. Results with alternative measures of horizontal inequality 20
Table 13. Countries included in subnational dataset of Education Inequality and Conflict 22
Table 14. Country representation in GED and EIC 22
Table 15. Descriptive statistics of subnational gap as absolute difference from the national mean 23
Table 16: Descriptive statistics on covariates in subnational regression models 25
Table 17. Results of logistic regressions with subnational unit disparity as a predictor of conflict in that subnational unit 26
Table 18. Geographic coverage in the Education Inequality and Conflict Dataset 32
Table 19. Regression results on final model (Model 4), with an alternative specification of decade bins 34
Table 20. Regression results: final model with robustness checks on length of time series and model specification 35
Table 21. Correlations between key variables in the global regression models (see Part I). 36
Table 22. Overview of regional variation from country mean 37

Executive Summary

Are countries where some ethnic or religious groups have systematically lower levels of education more likely to experience civil conflict than those where all groups have equal access to school? This is the central question in the growing literature investigating the relationship between horizontal inequalities (i.e., inequalities between ethnic, religious, and subnational groups) in education and violent conflict. This report takes a deeper look at this question, asking:

1. Does education inequality between ethnic and religious groups increase the likelihood of violent conflict?
2. Does education inequality between subnational regions within a country increase the likelihood of violent conflict in that country?
3. Does the relative disadvantage of a subnational region compared to the country as a whole increase the risk of violent conflict in that subnational region?

Methodology. We draw on two new datasets that offer substantially more comprehensive and fine-grained data on horizontal educational inequality than has previously been available– the Education Inequality and Conflict (EIC) Dataset, which spans five decades and includes data from nearly 100 countries, and the Subnational Education Inequality and Conflict Dataset (SEIC), which includes data on over 200 subnational regions in 24 nations in sub-Saharan Africa, from 1989-2012. In our analysis, the dependent variable is conflict onset, and the primary predictor of interest is education Group Gini – a measure of horizontal educational inequality in a given country or region and year, which are calculated from group differences in mean years of schooling. Having multiple observations for each country over time allows us to account for unobserved country-specific factors that may influence the likelihood of conflict in any one country. To carry out the analyses, we fit multilevel logistic regression models with random intercepts that take advantage of the longitudinal and clustered nature of the dataset.

Findings. We find a statistically significant and quantitatively large relationship between ethnic and religious inequality on likelihood of conflict in the 2000s, robust to multiple specifications of regression models. Specifically, we find that *one standard deviation in the Group Gini coefficient on mean years of education is associated with more than double odds of violent conflict*. However, this effect is not present across the entire historical period– in fact, while it comes out powerfully in the years since 2000, it is not present in the 1970-1990 period. In contrast, subnational educational inequality is a strong predictor of civil war regardless of the time period. *In terms of the relationship between a subnational region's relative inequality and its likelihood of conflict in sub-Saharan Africa, the results are inconclusive*. Findings suggest that subnational regions that are disadvantaged relative to the nation as a whole are more likely to experience conflict-related fatalities than are more advantaged regions. However, these findings are not robust to multiple specifications.

Overall, the findings show that in most recent years, countries with higher levels of horizontal inequalities in terms of mean years of schooling have been substantially more likely to experience violent conflict. While we acknowledge that the causality of this relationship cannot be established, we offer plausible explanations for the findings, including the increasingly severe implications of educational exclusion on individuals' life prospects, and suggest avenues for future research and data collection.

Introduction

This study is part of a research project commissioned by the UNICEF Peacebuilding, Education and Advocacy Programme (PBEA) Learning for Peace Initiative to examine the relationship between horizontal education inequality and violent conflict, and carried out by FHI 360's Education Policy and Data Center. For the purposes of this report, horizontal inequalities in education refer to inequalities in ethnic, religious and subnational groups' educational attainment, as measured by mean years of school.

Building on the literature, which has thus far found mixed support for the relationship between horizontal inequality in education and violent conflict, our analysis brings substantially more comprehensive and fine-grained data to the question of whether horizontal educational inequalities are associated with conflict (FHI 360, 2014).

This study examines three major research questions:

1. Does education inequality between ethnic and religious groups increase the likelihood of violent conflict?
2. Does education inequality between subnational regions within a country increase the likelihood of violent conflict in that country?
3. Does the relative disadvantage of a subnational region compared to the country as a whole increase the risk of violent conflict in that subnational region?

To answer these questions, the analysis draws on two newly created datasets – the Education Inequality and Conflict (EIC) Dataset, which spans five decades and includes data from nearly 100 countries, and the Subnational Education Inequality and Conflict Dataset (SEIC), which spans the years 1989-2012 and includes data on over 200 subnational regions in 24 nations in sub-Saharan Africa. The EIC dataset contains measures of inequality in the average educational attainment of young people (ages 15-24) from different ethnic and religious groups, as well as subnational regions, disaggregated by gender. It also includes information on the onset and incidence of civil conflict in country-year format.

The SEIC contains data at the level of the subnational (i.e., administrative) unit. It includes a measure of each subnational region's relative advantage or disadvantage, which is calculated as the difference in mean years of schooling between the subnational region and the national average, disaggregated by gender. It also includes the number of battle-related deaths annually in that region. The construction of both datasets is described in detail in the EIC Dataset Documentation, provided in the Technical Annex.

This report is structured into two parts: Part I draws on the EIC to answer the first two questions. It examines the relationship between horizontal inequalities at the national level and the likelihood that a country will experience conflict in the next five years.

To answer the third question, we require data on how educational opportunities vary within the same nation. As such, in Part II, we draw on the SEIC to examine how inequalities within the same nation affect different subnational regions' likelihood of experiencing conflict-related violence.

In both Part I and Part II, we first provide a descriptive overview of the measures of inequality and the indicators of conflict used in subsequent analyses and then conduct a series of logistic regression models examining the relationship between inequality and conflict. The report concludes with discussion and recommendations for future research.

Part I: Global analysis of group-based inequality

This research project examines the relationship between horizontal inequalities in education and the likelihood of violent conflict,[1] with the focus on horizontal inequalities in educational attainment of youth ages 15-24. In this section, we employ a global time series dataset covering 95 countries and 66 years. Our unit of observation is the country-year, with additional disaggregation by dimension of inequality and gender. The predictor variable is the level of horizontal educational inequality in a country in a given year– including inequality between ethnic groups, religious denominations, or primary subnational units. The outcome variable is a new conflict onset in a country at any point in the next five years, meaning the five years following the year in which the value of educational inequality is measured. Control variables include measures found to be associated with the likelihood of conflict in the literature, including democracy, anocracy, GDP per capita and prior conflict, also in country-year format. Regression analysis accounts for the binary nature of the outcome variable, as well as for the clustered nature of the panel dataset.

Dataset construction

The data for this analysis are drawn from the Education Inequality and Conflict (EIC) Dataset, which was constructed as part of this project. For detailed description of the dataset construction process, see Appendix B, Technical Annex. Measures of horizontal educational inequality were constructed as follows:

1. *Mapping identity groups*. Identity groups comprising 5% or more of the population were identified in source data (groups must have a common identity to be included, those falling in the "other" category are excluded);
2. *Data Extraction*. Group means of school attainment were extracted for each identity group, disaggregated by gender, in 10-year age increments, starting with the 15-24 age cohort;
3. *Back projection*. Back projections were applied to the extracted data from older age cohorts, in 10-year increments, to estimate educational attainment in previous decades. *This is done solely for data on ethnic and religious groups;* no back projection is applied to subnational data.
4. *Interpolation.* Education attainment values were interpolated in years without data or back projections; when interpolation created duplicate values due to overlapping time series, duplications were removed, keeping only the values from the most recent datasets;
5. *Calculation of inequality measures.* Group means and population weights were used to calculate country-level horizontal inequality measures, including the Group Gini coefficient, the Group Theil Index, a group-based coefficient of variance, and other measures;
6. *Merging with conflict data.* Education inequality data were merged with conflict data for analysis.

As noted above, we carry out back projections to estimate the educational attainment of each ethnic and religious group in previous decades. Using this method, the mean educational attainment of 15-24 year olds of a given ethnicity in the year 1975, for example, may be derived from the mean educational attainment of 35-44 year olds extracted in the year 1995, with an adjustment for differential mortality. Ethnic and religious groups are assumed to be stable over the years. By contrast, no back projection is

[1] Stewart (2000) defines horizontal inequality as inequality between identity-based groups (e.g., ethnic, religious, and subnational), which is distinct from vertical inequality, which is inequality between al individuals in a given country.

performed on subnational units, as their populations cannot be assumed to be the same over the course of several decades due to naturally occurring internal migration and changes in subnational boundaries.

Table 1. Number of observations by gender and dimension of inequality shows the number of observations by gender and identity dimension in the EIC dataset. In total, the dataset contains more than 16,000 observations (the exact number varies by gender and dimension of inequality); however, Table 1 also indicates that only 548 observations are available for measuring the effects of subnational inequality on conflict (Table 1). It is a clear that additional data are needed when examining subnational inequality, which we address in Part II. In this section, we focus on the country-year as the unit of analysis.

Table 1. Number of observations by gender and dimension of inequality

	Both	Male	Female	Total
Ethnic	2,483	2,466	2,539	7,488
Religious	2,803	2,778	2,812	8,393
Subnational	181	181	186	548
Total	5,467	5,425	5,537	16,429

Descriptive analysis

In this section, we describe the measure of horizontal inequality in education and the dependent variable, conflict onset, used in the analysis. The properties of the key variables used in our analysis are described below.

Horizontal Inequality in Education

For our global analysis, we use the *Group Gini (GGini) index* as our primary measure of horizontal educational inequality at the country level, following a suggested practice in the literature (Stewart, Brown and Mancini 2010). The index is based on the size of the differences between group averages within a given country, year, and type of inequality (i.e., ethnic, religious, and subnational) and the group's relative size as a proportion of the country's population.[2] While a separate GGini index was estimated for each level of education, we found that the distributional properties of *mean years of schooling* provide the optimal metric for examining education inequality. The GGini based on mean years of schooling can be interpreted as a measure of how concentrated the total stock of education is in any one ethnic or religious group. A GGINI of zero would mean that all ethnic and religious groups have the same mean years of schooling, while a GGINI of one can be understood loosely to correspond to a situation where one minority ethnic group has essentially exclusive access to all the education in the country, to the detriment of all other ethnic groups. Because it is a measure of concentration that accounts for the relative weight of each group in the population, it is inherently more sensitive to situations in which a minority has higher attainment than the majority.

[2] The construction of the index follows the formula below, where $\overline{y_r} = \frac{1}{n_r}\sum_i^{n_r} y_{ir}$ is group *r* mean value, *R* is the group *r*'s population size, p_r is group *r*'s population share, y_{tr} is the quantity of the variable of interest (e.g., income or years of education) of the i^{th} member of group *r*, Y_r is the value of *y* for group *r*, and *Y* is the grand total of variable *y* in the sample.

$$GGINI = \frac{1}{2\overline{y}}\sum_r^R\sum_s^S p_r\, p_s|\overline{y_r} - \overline{y_s}|$$

Dimensions of horizontal inequality. Our analysis examines three types, or dimensions of horizontal inequality – ethnic, religious and subnational, with separate GGini values estimates for each dimension. In measuring ethnic and religious inequality, we limit our analysis to countries with more than one ethnic and religious group, and establish a minimum cutoff, requiring groups to be at least 5% of the population. Horizontal inequality, unlike vertical inequality, by definition requires that a society be composed of more than one identity group. In our dataset, the GGini ranges from 0-0.965. However, the distribution is generally much tighter than the vertical educational GGini used by Bartucevicius (2014) and Ostby (2008), and has a substantial positive skew, with a particularly high outlier in ethnic inequality. Table 2. Summary statistics for Group GINI by group provides summary statistics of the GGINI by identity group. Most of the values fall between zero and 0.3 and a relatively small number outlier observations at the upper end of the distribution fall above 0.5.[3] This tighter distribution is expected, as our measure captures the differences between group mean values in the years of schooling, rather than the disparity between individuals.

Table 2. Summary statistics for Group GINI by group type

	Mean	SD	Min	Max
Ethnic	0.076	0.074	0	0.965
Religious	0.064	0.076	0	0.528
Subnational	0.098	0.09	0	0.578

Nonetheless, because we are measuring inequality using mean years of schooling for identity groups and regions as a whole, even a small difference in horizontal inequality can mean real differences in the life opportunities of members of different groups. A one year difference in mean years may translate into the difference between graduating high school, and receiving the concomitant benefits, and not graduating.

Figure 1 shows the distribution of the GGini for mean years of schooling by identity group type. As the graph indicates, inequality is generally higher between geographic subnational units than it is for the identity-based groups, religion and ethnicity. This is generally true in all world regions, as shown in Figure 2, with the exception of Eastern Europe where ethnic inequality is highest.

[3] This is in contrast to the commonly used *Gini* index of wealth, which is considered "low" at 0.3 and below, and "high" at 0.6 and above.

Figure 1. Distribution of horizontal inequality in education by type

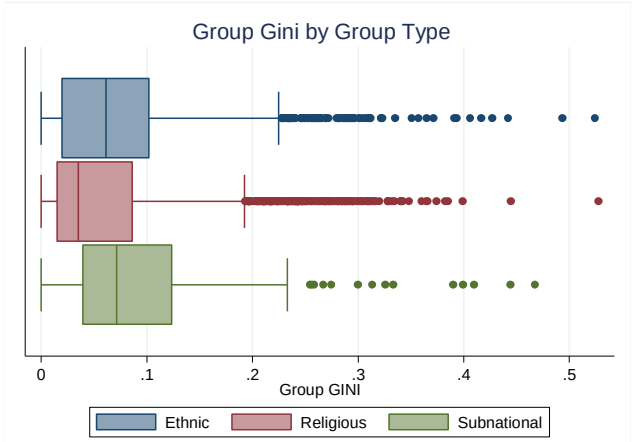
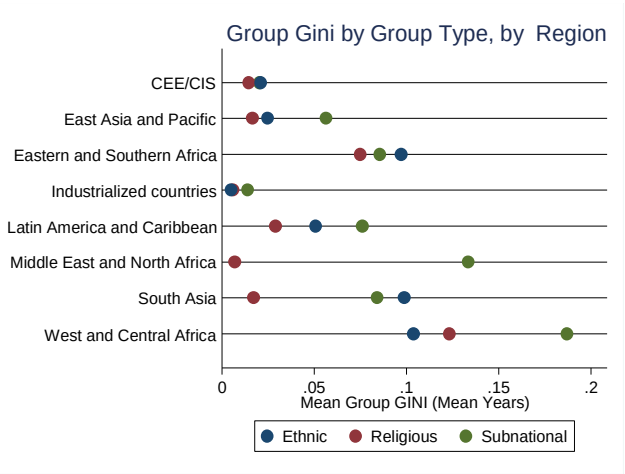


Figure 2. Mean group GINI by group type and world region



Gender. Our measure of inequality differentiates inequality by gender, separately measuring educational disparities between males of different identity groups and females of different groups. Across the board, inequalities between women are larger than those between men, with somewhat wider gaps along the ethnic dimension (Figure 3). However, the gender-disaggregated GGini indices are highly correlated, which indicates that where inequality is high in one gender, it tends to also be high in the other. This is an important finding that has implications for our regression analysis, as it suggests that results are unlikely to be different for males and females.[4]

Figure 3. Mean Group GINI by gender and group type

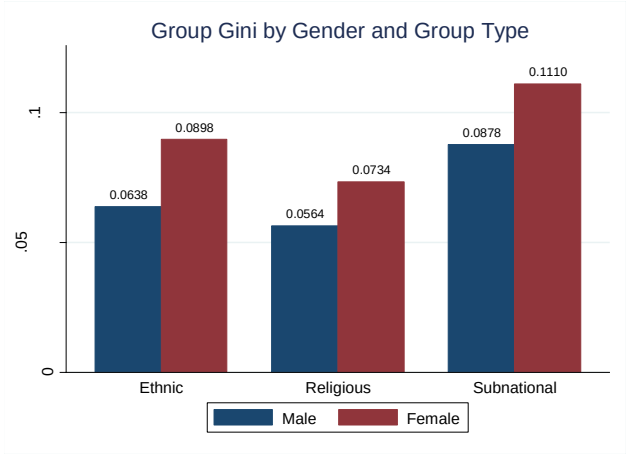


Table 3. Correlation of inequalities between males and females, by group type

Identity Group Type	Male-Female Correlation
Ethnic	0.85
Religious	0.88
Subnational	0.91

Downward trend of inequality in education. Around the world, access to education has increased dramatically over the last five decades. As enrollments in education systems grew, the stock of human

[4] We had hypothesized that inequality between males will be a stronger predictor of violent conflict than inequality between females.

capital, measured in years of schooling, became more equitably distributed. This is because unlike income, which has no ceiling, there is a natural maximum number of years of schooling one can attain in every educational system (i.e., the total duration of schooling). Therefore, as more individuals gain access to the mass education system, education becomes less concentrated in any one subgroup. As a result, we find that over time, horizontal inequalities in education have declined in every region of the world (Figure 4, Figure 5).

The most dramatic declines in horizontal inequalities occurred in countries with the highest horizontal inequalities in the 1960s, particularly in sub-Saharan Africa. The regional mean in sub-Saharan Africa decreased by roughly half, from above 0.17 in 1960 to 0.08 in the 2000s. Horizontal inequalities across religious groups has also declined in the Middle East and North Africa region. For the other world regions, horizontal inequalities in religion have always been relatively small, and remain so. The presence of a time trend in horizontal inequality suggests the importance of controlling for time in our subsequent regression analysis.

Figure 4 . Ethnic inequality in education across time (MENA not available)

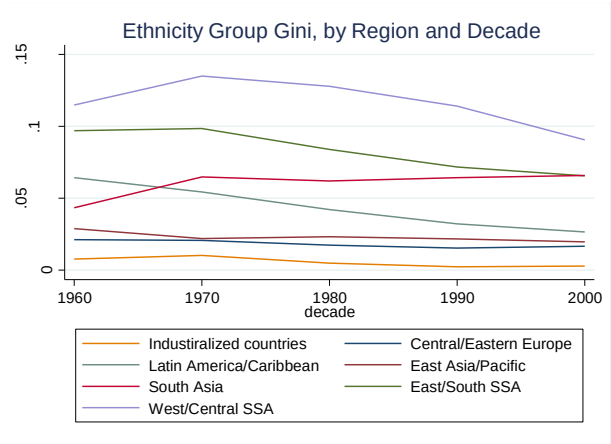
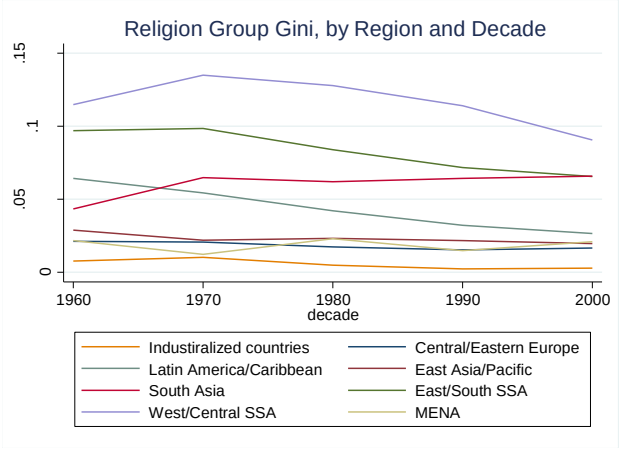


Figure 5. Religious inequality in education across time



Conflict Onset

Our measure of violent conflict, ***conflict onset***, is borrowed from the Uppsala Conflict Data Program (UCDP) datasets. Specifically, for our global analysis we use the onset variable from the UCDP Onset of Intrastate Armed Conflict, which spans 66 years (1946-2011), and includes annual observations on conflict onset in over 180 nations (Themnér and Wallensteen 2012). In the UCDP Onset dataset where conflict is defined precisely as at least 25 battle-related deaths in one calendar year and onset means a new outbreak after a period of peace.[5] To supplement the dataset with the most recent available data, we coded conflict onsets for 2011-2013 using UCDP definitions. For the subnational-level analysis, we use

[5] UCDP defines armed conflict as follows: "an armed conflict is a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths in one calendar year" (UCDP 2014). Because it does not capture instances of conflict between two non-state actors, the measure of conflict may underestimate the extent of ethnic or religiously motivated conflict around the world.

the Uppsala Geo-Referenced Event Dataset (GED), which provides geographic location of conflict events for sub-Saharan Africa for 1989-2010 (see Part II below).

As is common in the literature, we adopt a definition of conflict onset that includes a two-year lag: incidence is coded as new onset if at least two years have passed since the last observation of the conflict. This definition is widely used in the literature on conflict; however, it also may introduce artificiality to the idea of onset in the case of protracted conflicts. In particular, given the accounting of battle deaths by calendar year, it is possible that an incidence of conflict that spanned the New Year would not be recorded and then would enter the dataset as a new onset of conflict, when in fact it may actually be simply the continuation of an existing conflict.

Conflict around the world. As Table 4 shows, in total, our dataset includes 95 countries with mean years of schooling and important covariates, of which 57 different countries experience a new conflict onset and 63, equal to roughly two-thirds (66.32%), experience a conflict at some point in the time period, a rate quite a bit higher than the global mean (51.67%). We also find that while our dataset replicates regional percentages well in some regions, namely North America, Eastern Europe and Africa, it tends to over-represent conflict affected countries in Asia and the Middle East and North Africa, while underrepresenting conflict affected countries in Latin America and the Caribbean. Although it would be preferable for the dataset to more closely mirror rates of global conflict onset, the EIC is limited by availability of data on educational attainment. It remains the most comprehensive dataset available to date on educational inequalities worldwide.

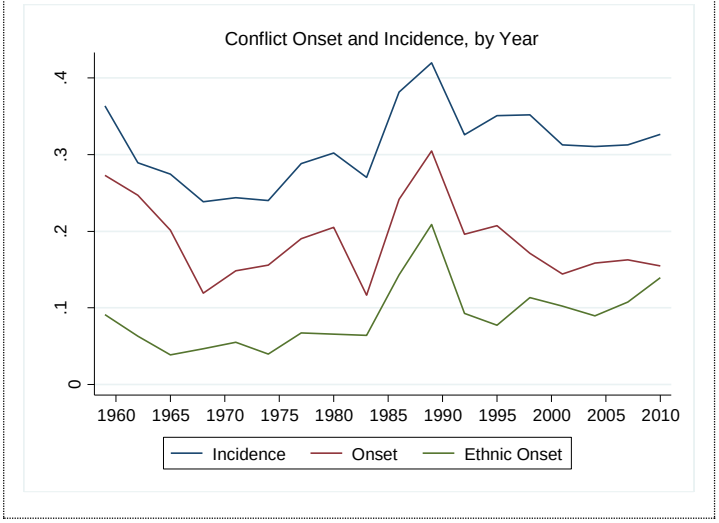
Table 4. Geographic coverage and conflict incidence in the UCDP and EIC, by world region

	World (Source: UCDP)			EIC dataset		
World Region	Number of countries	Ever In conflict	% in Conflict	Number of countries	Ever In conflict	% in Conflict
North America and Western Europe	21	4	19.0%	6	1	16.67%
Eastern Europe	29	11	37.9%	12	5	41.67%
Latin America and Caribbean	22	17	77.3%	22	14	63.64%
Africa	41	34	82.9%	36	29	80.56%
Asia	22	14	63.6%	15	11	73.33%
Middle East and North Africa	19	13	68.4%	4	3	75.00%
Total	180	93	51.67%	95	63	66.32%

Note: World Bank regions used in Table 4

Conflict onset over time. Prior research on civil conflict has pointed out that around the world, the nature of conflict has shifted over the past four decades from primarily inter-state to intra-state, or civil conflicts. Our dataset, while not capturing every country in the world, reflects global trends in conflict outbreak and incidence. As Figure 6. Conflict onset and incidence, by year shows, new onsets of conflict went on a slight downward trend between 1960 through mid-1970's, before rising precipitously in the late 1980s and early 1990s, and finally returning to pre-1980 levels. As such, there does not appear to be a time trend in either direction, but rather the outbreaks of new conflicts follow an up and down trajectory.

Figure 6. Conflict onset and incidence, by year



It is important to note that the sample of countries for which data on educational inequality are available includes a higher proportion of countries experiencing conflict than does the world as a whole (Table 5). This means that the relationship between education inequality and conflict will be estimated with a slight skew towards conflict-prone countries. However, this oversampling appears to be consistent over time, and we do not believe it is systematically linked to countries with higher rates of ethnic, religious or subnational horizontal inequality.

Table 5. Country coverage and conflict incidence in UCDP and EIC, by decade

	World (Source: UCDP)				EIC dataset	
Decade	# of countries	Countries in conflict	% in Conflict	# of countries	Countries in conflict	% in Conflict
1960s	136	35	25.7%	83	27	32.53%
1970s	150	42	28.0%	90	31	34.44%
1980s	151	51	33.8%	91	44	48.35%
1990s	173	61	35.3%	89	41	46.07%
2000s	173	48	27.7%	86	27	31.40%

In the next section, we describe the methodology and results of the regression analysis predicting new conflict onset as a function of horizontal inequality.

Regression analysis: Horizontal Inequality and Conflict

Prior to fitting regression models predicting conflict onset, we modify our key variables as follows:

- Conflict onset is transformed into a continuous time series where for each country-year observation, 1 denotes the presence of new conflict onset in the following five-year period, and 0 denotes continuous peace, if no conflict was experienced. Years of continuing conflict, if spanning the entire five-year period, are set to missing.
- Consequently, the time series for horizontal education inequality measures, as well as other covariates, are truncated at 2008 or earlier, to allow for the five-year lag between the measurement of inequality and the measurement of conflict onset.

Covariates

In addition to the outcome variable, conflict onset, and key predictor, horizontal education inequality, we include a number of relevant covariates in our regression analysis. These variables have been shown in prior research to be strongly associated with conflict occurrence, and may therefore improve the precision of our models by parsing out the variance related to educational inequality from the variance related to other factors. It is important to remember that we do not seek to explain the variation in conflict onset itself, but rather to identify whether a link exists between horizontal education inequality and the likelihood of violent conflict breaking out in the immediate future. Important control variables include:

- *Level of economic development*. Prior research (Hegre & Sambanis 2006; Montalvo & Reynal-Querol 2005, Brown 2009) has found that countries with lower levels of economic development are associated with higher rates of conflict. We use a covariate for gross domestic product (GDP) per capita, logged. GDP per capita is taken from the Penn World Tables, which has the most complete data for the countries in our analysis over the time period.
- *Past history of conflict.* We calculate a variable measuring peace years, or the number of years that have passed since the last incidence of conflict, based on UCDP data.
- *Political regime.* Research has found that democracies and anocracies are both more likely to experience conflict that authoritarian states (see, for example, Vreeland 2008, also Hegre & Sambanis 2006; Brown 2009; Hegre et al 2001). As such, we also control for democracy and anocracy, operationalized as binary variables, drawn from Polity IV dataset.
- *Population.* To control for a country's size, we include a measure of total population, logged, from the World Development Indicators. This is in line with previous literature (Collier & Hoeffler, 2004; Fearon & Laitin, 2003; Hegre & Sambanis 2006).
- *Geographic terrain.* Fearon and Laitin (2003) found that countries with more mountainous terrain are more likely to experience insurgencies. To control for geographic terrain, we use Fearon and Laitin's estimates of mountainous terrain in a country.
- *Ethnic and religious fractionalization.* A number of prior studies (Fearon & Laitin 2003; Hegre & Sambanis 2006) have used controls for diversity, on the hypothesis that countries with more socio-politically relevant groups will be more likely to experience conflict. We proxy a measure of ethnic and religious fractionalization by including the number of groups over 5% of the population, as calculated from the EIC.
- *Economic inequality*. Although this study is focused on horizontal inequality, prior research (see Bartucevicius 2014) finds that vertical economic inequality is also associated with conflict. We also control for vertical inequality with a wealth GINI index.

Table 6 shows the basic statistics for each of the covariates included in the model.

Correlations. Prior to fitting regression models, we examine how our measure of horizontal inequality and covariates are correlated with each other with important covariates identified in the literature. Table 20, in the Appendix, shows correlations between our measure of horizontal inequality (GGINI) and other covariates, including: GDP per capita, population, democracy, vertical educational inequality, wealth inequality and the percentage of the country that is mountainous terrain. We found that a number of variables were highly correlated, which may cause problems associated with multicollinearity if jointly

included in regression models. To avoid problems, we limit our analyses to a select number of key covariates.

Table 6. Descriptive statistics for variables included in regression models

Covariate	Mean	SD	Min	Max	Observations	Source
Group GINI	-0.09	0.85	-0.96	3.58	3427	EPDC EIC
Year	1984.67	12.72	1960	2008	3427	--
GDP per capita (logged)	6.95	1.27	3.91	10.82	2892	Penn World Tables
Peace Years	15.53	14.78	0	63	3123	UCDP
Population (logged)	15.89	1.51	11.44	20.75	3403	WDI
Youth Population (% Total)	26.20	2.22	18.28	33.27	3427	UNPD
Democracy	0.30	0.46	0	1	2986	Polity IV
Anocracy	0.28	0.45	0	1	2986	Polity IV
Number of Groups	3.77	1.79	2	9	3427	EPDC EIC
Wealth GINI Index	44.32	10.45	15.50	78.60	2408	UN-WIID
Mountain Terrain (%, logged)	2.22	1.48	0	4.42	3010	Fearon and Laitin
Oil and Gas Production (logged)	1.85	2.49	0	9.44	3136	Ross
Education Spending	13.36	7.49	2.27	58.16	206	WDI
Educational Attainment (Years)	6.25	2.92	1.34	12.84	3427	EPDC EIC

Results

Given that our dataset for this part of the analysis is clustered by country, we fit a series of models that account for the grouped nature of the data and the inter-dependence of error terms within each country panel. Initially, we fit models for ethnic and religious inequality only, since they have a substantially larger number of observations. We then follow these models with examination of the effects of subnational inequality, which has different country and year coverage, given that no back projection was performed on educational attainment data.

Ethnic and Religious Inequality

The EIC calculates separate indicators of horizontal inequalities for ethnic and religious groups. However, in many countries, we have only one value – either ethnic or religious. Therefore, for the purpose of the regression analyses, we create a combined dataset that draws on either ethnic or religious horizontal inequality, whichever is available. We prioritize ethnically based inequalities because the descriptive analysis above suggests that they are larger worldwide than are religious inequalities. As such, the combined dataset includes an indicator of ethnic horizontal inequality if present, and if not present, an indicator of horizontal inequalities across religious groups. This allows us to capitalize on the breadth of our dataset and ensure as many countries as possible are included in the analysis. We subsequently disaggregate by type of inequality and by gender; however, we do not find statistically significant differences in the likelihood of conflict than with the combined model.

Table 8 presents the results from the analysis of the relationship between ethnic and/or religious inequality on violent conflict. In Models 1-4 conflict onset is regressed on the combined dataset, which is ***ethnic OR religious*** inequality. Model 4 is the most inclusive model, as it accounts for the most important covariates while also drawing on the full dataset. Models 5-6 then distinguish between ethnic and religious inequality, and Models 7-8 disaggregate by gender. Table 7 provides brief model descriptions.

Table 7. Model specifications

Model #	Specification	Description
1	$Y = \beta_0 + \beta_1 GGiniER + \beta_2 T_i + \beta_3 X_i + \varepsilon_j$	Logistic regression with clustered standard errors and controls, time T in years
2	$Y = \beta_{0j} + \beta_1 GGiniER + \beta_2 T_i + \beta_3 X_i + \varepsilon_j$ $\beta_{0j} = \gamma_{00} + R_{0j}$	Random intercept model with basic controls, time in years (In notation to the left, the intercept β_{0j} consists of a fixed portion γ_{00} and random portion R_{0j}, set at country level)
3	$Y = \beta_{0j} + \beta_1 GGiniER + \beta_2 D_i + \beta_3 GGini_i \times D_i + \beta_4 X_i + \varepsilon_j$ $\beta_{0j} = \gamma_{00} + R_{0j}$	Random intercept model with basic controls, time D_i in decades (2000's is reference category), time interaction effect $GGini_i \times D_i$, and basic controls (random part of the intercept shown as above)
4	$Y = \beta_{0j} + \beta_1 GGiniER + \beta_2 D_i + \beta_3 GGini_i \times D_i + \beta_4 X_i + \beta_5 Z_i + \varepsilon_j$ $\beta_{0j} = \gamma_{00} + R_{0j}$	Random intercept model, time D_i in decades (2000's is reference category), time interaction effect $GGini_i \times D_i$, basic controls and additional covariates Z_i
5-6	$Y = \beta_{0j} + \beta_1 GGiniER + \beta_2 D_i + \beta_3 GGini_i \times D_i + \beta_4 X_i + \beta_5 Z_i + \varepsilon_j$ $\beta_{0j} = \gamma_{00} + R_{0j}$	Same as Model 4, but specified separately for ethnic and religious inequality
7-8	$Y = \beta_{0j} + \beta_1 GGiniER + \beta_2 D_i + \beta_3 GGini_i \times D_i + \beta_4 X_i + \beta_5 Z_i + \varepsilon_j$ $\beta_{0j} = \gamma_{00} + R_{0j}$	Same as Model 4, but specified separately for male and female inequality (ethnic and religious combined)
9	$Y = \beta_0 + \beta_1 GGiniS + \beta_2 T_i + \beta_3 X_i + \varepsilon_j$	Logistic regression with clustered standard errors and controls, time T in years (Same as # 1, but for Subnational)
10	$Y = \beta_{0j} + \beta_1 GGiniS + \beta_2 T_i + \beta_3 X_i + \varepsilon_j$ $\beta_{0j} = \gamma_{00} + R_{0j}$	Random intercept model with basic controls, time in years Subnational inequality $GGiniS$
11-12	$Y = \beta_{0j} + \beta_1 GGiniS + \beta_2 D_i + \beta_3 GGini_i \times D_i + \beta_4 X_i + \beta_5 Z_i + \varepsilon_j$ $\beta_{0j} = \gamma_{00} + R_{0j}$	Random intercept model, time D_i in decades (2000's is reference category), interaction effect $GGini_i S \times D_i$, basic controls and regime covariates $Z_i$in Model 12
13-14	$Y = \beta_{0j} + \beta_1 GGiniS + \beta_2 D_i + \beta_3 GGini_i \times D_i + \beta_4 X_i + \beta_5 Z_i + \varepsilon_j$ $\beta_{0j} = \gamma_{00} + R_{0j}$	Same as Model 12, but specified separately for each gender (Male and Female)

As Table 8 shows, in Model 1 we begin by fitting conflict onset on horizontal inequality with simple controls for GDP per capita, peace years and the year of observation, which is centered at 1985, with robust standard errors clustered at the country level. We include a control for historical year, because we know that horizontal inequalities have been decreasing over time, as access to schooling has increased, and that the likelihood of conflict onset has changed over time in response to larger macro-political changes. This simple model suggests that overall, horizontal inequality has had little to no effect on the likelihood of conflict onset.

However, we anticipate that countries will have varying propensities to experience conflict based on unobserved factors, which are not controlled for in basic logistic regression models. To control for unobserved country differences that remain stable over time, in Model 2, we fit a random intercepts model. In Model 2, we find support for previous findings in the literature – countries with higher GDP per capita are less likely to experience a new conflict onset. A comparison of model fit between Models 1 and 2, examining the Bayes Information Criteria (BIC) show that Model 2 is a significantly better fit, suggesting these covariates improve the model. Model 2, however, also shows no statistically significant relationship between horizontal inequality and conflict onset.

Although Model 2 is a better fit than Model 1, the descriptive analyses above suggest that the relationship between time and conflict is not linear, but rather, that *countries' propensity for conflict is different in every decade*. Therefore, in Model 3, we include binary variables for each decade, and interact

our measure of horizontal inequality with each of these decades. The reference decade is the 2000s, which is the most recent decade and also the one for which we do not have to conduct back projections, meaning it requires the fewest assumptions about change over time. As with prior models, we include basic controls for GDP per capita, and years since last conflict (i.e., peace years). As Model 3 shows, horizontal inequality is strongly, positively associated with conflict onset in the 2000s, and generally less correlated with conflict onset in preceding decades.

In Model 4, we include additional covariates suggested by the literature on conflict, namely population size (logged), democracy, anocracy and a proxy for ethnolinguistic fractionalization. Our findings are consistent with those in prior studies – both democracy and anocracy are consistently positively associated with onset and both are statistically significant. Similarly, the association between anocracy and conflict is higher than is the association between democracy and conflict. Population is positively correlated with onset and is consistently statistically significant. Our measure of ethnic and religious fractionalization is not statistically different from zero, suggesting it has little effect on conflict onset.

Importantly, even after controlling for these important covariates, we still find that in the 2000s, *higher horizontal inequality is positively associated with conflict onset.* Model 4 shows that in the 2000s, *a one standard deviation increase in horizontal inequality in educational attainment more than doubles the odds that a country will experience a conflict in the next five years*. The relationship between inequality and conflict is significantly lower in earlier decades – again suggesting the effect is most pronounced in the most recent era.

Model 4 is our preferred model. This model ensures the strongest level of statistical power by pooling ethnic and religious inequality measures, and captures inequality for the entire population, irrespective of gender. It also allows for separate fixed effects on the different time periods, making the model more informative as to the likely changes in the relationship between our measure of education inequality and conflict onset depending on the time period in question. We find the strongest effects during 2000s, which coincides with the greatest access to education and the lowest levels of inequality. This may suggest horizontal inequality is more consequential in recent years, where access to basic education is available to all but the most marginalized populations. We examined our dataset carefully and were able to verify that observations from the year 2000 are not substantially different from the rest of the dataset in terms of country coverage and the types of countries that were included. Additionally, we also ran a series sensitivity checks (below) that show that this finding is robust to alternative specifications.

Models 5-8 expand on Model 4, by disaggregating our predictor variable by inequality type (ethnic and religious, Models 6-7) and gender (Models 8-9). The results from Models 5-8 are similar to Model 4. The estimate on ethnic inequality is stronger than on religious inequality; however, this may have to do with country coverage in each of these models, as indicators of religious and ethnic identity was not available for all countries in our dataset. The estimate for horizontal inequality among females is higher than that among males, suggesting that it is inequality between females that is likely driving up the effect associated with horizontal inequality in general.

Table 8. Regression estimates: Ethnic and Religious Inequality

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Model	Logit	Random Effects	Random Effects	Random Effects	Random Effects	Random Effects	Random Effects	Random Effects
Group Identity	Ethnic or Religious	Ethnic or Religious	Ethnic or Religious	Ethnic or Religious	Ethnic	Religious	Ethnic or Religious	Ethnic or Religious
Gender	Both	Both	Both	Both	Both	Both	Male	Female
GGINI: Horizontal Inequality	0.889	1.127	3.092***	2.751**	2.805*	3.291**	2.413*	3.220***
	0.12	0.18	0.85	0.96	1.26	1.35	0.99	0.89
Year	1.003	0.998						
	0.01	0.01						
1990s			1.439+	2.097**	1.5	1.408	1.479	2.979***
			0.3	0.57	0.47	0.43	0.4	0.88
1980s			1.936**	4.478***	2.082+	4.059***	2.509*	7.329***
			0.44	1.65	0.9	1.66	0.94	2.82
1970s			1.115	4.124**	0.675	5.118**	1.738	5.143**
			0.33	2.03	0.41	2.74	0.88	2.64
1960s			1.098	3.912*	0.517	5.441*	1.641	6.811**
			0.43	2.48	0.39	3.7	1.06	4.43
1990s # Group GINI			0.320***	0.219***	0.193***	0.375**	0.285***	0.297***
			0.08	0.07	0.07	0.12	0.1	0.07
1980s # Group GINI			0.228***	0.113***	0.063***	0.262***	0.135***	0.176***
			0.06	0.04	0.03	0.09	0.05	0.05
1970s # Group GINI			0.241***	0.235***	0.226***	0.316**	0.267**	0.183***
			0.07	0.08	0.1	0.11	0.11	0.06
1960s # Group GINI			0.462*	0.505+	0.447+	0.491+	0.531	0.241***
			0.14	0.2	0.21	0.19	0.24	0.1
GDP per capita (logged)	0.89	0.997	0.869	0.841	0.306***	1.203	0.659+	0.99
	0.13	0.16	0.15	0.19	0.1	0.31	0.16	0.24
Peace Years	0.893***	0.882***	0.884***	0.873***	0.885***	0.891***	0.876***	0.877***
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Peace Years Squared	1.002***	1.004***	1.004***	1.004***	1.004***	1.003***	1.004***	1.004***
	0	0	0	0	0	0	0	0
Population (logged)				2.145*	2.182*	2.026*	2.091*	1.976*
				0.65	0.79	0.65	0.63	0.58
Democracy (0/1)				2.324**	2.209*	2.246*	2.094*	2.038*
				0.73	0.82	0.89	0.65	0.63
Anocracy (0/1)				2.853***	2.205**	3.521***	2.989***	2.905***
				0.7	0.63	1	0.74	0.71
Ethnic Groups				1.152	1.117	0.721+	0.869	0.955
				0.19	0.21	0.14	0.14	0.1
Wealth Inequality (GINI)				1.024	1.014	1.032+	1.031+	1.033+
				0.02	0.02	0.02	0.02	0.02
Constant	0.003	1.939	0.083***	0.002***	0.012**	0.009***	0.008***	0.002***
	0.07	35.04	0.04	0	0.02	0.01	0.01	0
Random Effects Parameters								
S.D. of Constant		3.327***	2.762***	3.319***	3.044***	3.132***	3.264***	3.356***
S.E.		0.5	0.43	0.59	0.62	0.64	0.59	0.6
N	2648	2789	2789	1928	1339	1516	1894	1922
N. Countries	2412.128	2137.097	2142.496	1466.585	1056.722	1218.648	1431.665	1479.137
BIC	0.889	1.127	3.092***	2.751**	2.805*	3.291**	2.413*	3.220***

Notation: *p<0.05, **p<.01; ***p<.001

Inequality between subnational regions and likelihood of conflict

As Table 2 shows, the magnitude of inequality between subnational regions is greater than that of ethnic and religious inequality. In this section, we test whether this means that inequality between regions has a stronger effect on the country's likelihood of experiencing violent civil conflict.

We run our models with subnational inequality as a predictor of violence separately because of the different composition of the dataset. As noted in the documentation and in Dataset section above, no back projection or interpolation was performed for subnational educational attainment, and hence the data on horizontal subnational educational inequality are only available for those countries and years for which we have actual data from surveys and censuses. This is because we felt the assumptions of unchanging group composition were too strong in the case of subnational unit population to be plausible for back projection, and because we knew that a number of administrative restructuring efforts had taken place in many countries where the boundaries of subnational regions have changed, making it difficult to back project from present-day data.[6]

In this section, we present an analysis of *subnational inequality as a dimension*, albeit with a smaller dataset that cannot fully account for all the covariates (see sample size in Table 1 above). Here, we examine the likelihood of conflict in the country as a whole, with that country's level of between-region inequality GGini as a predictor. In Part II, we take this analysis further and examine the issue differently, by placing conflict at the level of the subnational unit itself, and using the disparity between the unit and the national mean to predict conflict. Therefore, the principal difference between what is shown here and what is shown in Part II is the location of conflict (country-level vs subnational), and the conceptualization of inequality (between all regions vs. the region vs the national mean).

for the earliest observations).

Table 9 shows the results of the logistic models. As done previously, Model 9 shows a simple logistic regression model with clustered standard errors, while Models 10-14 are multilevel panel logistic models that fit random intercepts for each country. As for the earliest observations).

Table 9 shows, the odds ratios of conflict at the country level associated with between-region inequality in education are quite similar in magnitude to the results we saw for ethnic and religious inequality in Table 8, with an important exception: Models 9 and 10 both show a statistically significant effect of inequality (an odds ratio of 2:1 for conflict onset in countries with horizontal inequality that is one standard deviation above the mean). Unlike the models above, we find that the main predictor for horizontal inequality (GGINI) are significantly associated with conflict onset in both the logistic and random effects models that include only a simple control for year. Model 10 shows that a country with a GGINI index that is one standard deviation higher (roughly 0.09), has 60% higher odds of experiencing conflict than one with the mean GGINI score (roughly 0.10).

Model 11 includes decade interactions and Model 12 includes important covariates. Models 13 and 14 test each gender separately. In the models with decade interactions, we do find similar trends – that the

[6] Vertical alignment of subnational regions was, however, performed within a separate dataset for Africa, and analysis of subnational conflict likelihood is presented in the following section.

coefficient on horizontal inequality is highest in the 2000s; however, the interaction terms are not significant in other decades. It is possible that this is due to the small sample size.

These models show, despite their small sample sizes, that *the same high and statistically significant odds ratios associated with 1 standard deviation increase in between-regions subnational inequality: the difference in odds of conflict between a relatively equal and a highly unequal country is placed somewhere between 3:1 and 4:1*. This is also true for the decade of the 2000's, but with this specification of inequality, *the effect holds across the entire time series* (which is truncated at 1970 for the earliest observations).

Table 9. Results of logistic regressions with Subnational inequality as a predictor

	Model 9	Model 10	Model 11	Model 12	Model 13	Model 14
Model	Logit	Random Intercepts	Random Intercepts	Random Intercepts	Random Intercepts	Random Intercepts
Group Identity	Subnational	Subnational	Subnational	Subnational	Subnational	Subnational
Gender	Both	Both	Both	Both	Male	Female
GGINI: Horizontal Inequality	1.906***	1.591*	3.076*	3.143*	3.429*	2.649*
	0.37	0.3	1.45	1.49	1.81	1.03
Year	0.987	0.971				
	0.03	0.02				
1990s			1.88	1.893	2.081	1.698
			1.43	1.47	1.56	1.36
1980s			2.858	3.165	3.513	3.438
			2.67	3.18	3.55	3.32
1970s			5.668+	7.026+	7.742+	7.510+
			5.86	7.55	8.73	8.13
1990s # Group GINI			0.582	0.584	0.48	0.688
			0.27	0.27	0.26	0.26
1980s # Group GINI			0.489	0.501	0.437	0.571
			0.26	0.27	0.28	0.24
1970s # Group GINI			0.125+	0.119+	0.118+	0.136+
			0.14	0.14	0.13	0.16
GDP per capita (logged)	0.791	0.838	0.747	0.755	0.755	0.775
	0.29	0.25	0.31	0.32	0.33	0.34
Peace Years	0.955+	0.948*	0.959+	0.958+	0.953*	0.961+
	0.02	0.02	0.02	0.02	0.02	0.02
Population (logged)	1.861**	1.954**	2.011**	1.938**	1.977**	1.920**
	0.37	0.41	0.45	0.46	0.47	0.47
Democracy (0/1)				1.248	1.234	1.476
				1	1	1.19
Anocracy (0/1)				2.065	2.079	2.656
				1.57	1.57	2
Constant	0.571	0.378+	0.100**	0.069**	0.081**	0.048**
	0.35	0.2	0.07	0.07	0.08	0.05
Random Effects Parameters						
S.D. of Constant		0	0.565	0.698	0.407	0.976
S.E.		0.4	0.64	0.64	0.69	0.68
N	153	158	158	153	153	157
BIC	152.395	149.035	168.086	176.21	177.039	177.433

Interpretation of Results

These results suggest that ***ethnic and religious inequality in education (as measured by mean years of schooling) is indeed predictive of violent conflict, and this is true for both ethnic and religiously-based inequalities.*** We find that differences of one standard deviation on the inequality variable, which in the 2000s translates roughly into an increase in the GGINI from 0.054 to 0.101, would more than triple the odds of violent conflict taking place. Further, we find that ***subnational inequality is also strongly associated with higher odds of violent conflict***, and this relationship is present throughout our time series.

Table 10. Marginal probability of conflict onset at different levels of horizontal inequality (ethnic or religious) shows the marginal probabilities of conflict in each decade, calculated from our preferred model (Model 4), which uses ethnic and religious inequality as the predictor.

Table 10. Marginal probability of conflict onset at different levels of horizontal inequality (ethnic or religious)

Decade	Probability of Conflict at 1 SD Below Mean	Probability of Conflict at Mean Inequality	Probability of Conflict at 1 SD Above Mean	Probability of Conflict at 2 SD Above Mean
1960	0.124	0.150	0.179	0.213
1970	0.195	0.154	0.120	0.093
1980	0.291	0.161	0.082	0.039
1990	0.140	0.105	0.077	0.056
2000	0.034	0.066	0.123	0.213

Robustness Checks

In the next set of models, we examine the findings from prior models to determine whether they are robust across a variety of specifications. In this section, we provide the results of these robustness checks, displaying the odds ratios associated with one standard deviation increase in horizontal inequality, with interaction terms for time period (the most recent decade being the reference category).

Alternative dataset specifications

We verify the robustness of our Group Gini estimate by altering the length and coverage of our dataset, to test for selection bias and measurement error associated with the construction of key variables. Table 11 presents the estimates of odds ratios obtained on the Group Gini of mean years of education in these alternative specifications. The full output of these models is presented in Appendix B: Sensitivity Checks.

Table 11. Results of robustness checks: alternative specifications of data

Alternative model specification	Inequality effect	Model Number
Alternative Decades (decade bins specified differently)	2.472 **	15
No Long Projections: back projections no longer than 20 years)	2.074 +	16
Five Decade Only (dropped 1960's)	3.079 ***	17
Fixed Effects	2.435 **	18
Note: Cell entries are odds ratios of conflict between countries one standard deviation higher than the mean on horizontal inequality, over countries exactly at the mean. Other variables and grouping levels same as in Model 4 above.		

Model 15 changes year bins to test for robustness across different periods of time. Rather than using decade bins, we divide the time series in the middle of each decade: 1960-1974, 1975-1985, 1986-1995 and 1996-2008, our last year of observation. Model 16 uses the traditional decade definition, but removes all observations from long projections (20+ years), to control for the possibility that bias was introduced in the back projection calculations. Model 17 includes back projections, but is limited to countries for which we have data spanning five decades. This assures that the findings are not driven by

countries leaving or entering the dataset in the 2000s. Finally, Model 18 shows the relationship between inequality and conflict occurrence using *fixed effects*, which controls for time-invariant country-specific characteristics. This model offers a slightly different interpretation than the random effects models used previously, namely, as inequality in a single country changes, how does its likelihood of conflict change?

Across these models, we find that our estimate is somewhat sensitive to the length of the time series, but the effect fluctuates between an odds ratio of 2.4 and 3.1, and remains statistically significant.

Alternative measures of inequality

Additional models examine an alternate specification of inequality. We run our model with three alternative measures: the Group Theil Index, a modified version of the Lineq2 Index (Cederman et al., 2011), and Group Coefficient of Variance (Stewart et al., 2010; Mancini, Stewart and Brown, 2008). The Theil Index captures the population-weighted ratios of the group mean to the national average for educational attainment, summing them up by dimension of inequality. The Lineq2 Index offers a slightly different interpretation, as it is based on the ratio of the highest group's mean attainment to the lowest group's mean attainment, regardless of the group population weight.[7] Finally, the coefficient of group-level variance (G-COV) is a measure of dispersion, and constructed as a sum of squared absolute deviations from the mean multiplied by population weight, by group type.[8]

Table 12. Results with alternative measures of horizontal inequality

Measure of Inequality	Inequality Effect
Group Theil Index	1.78 **
Modified Lineq2 Index	3.89 **
Coefficient of Group-level Variance (G-COV)	1.73 **

Because these measures capture inequality slightly differently, we present the synopsis of odds ratios of conflict associated with a one-standard deviation change in inequality. The coefficients on horizontal inequality in the most recent period are presented in Table 12. Full results are available upon request. In each of the models, we find that in the most recent decade – the 2000s – higher horizontal inequalities are positively correlated with conflict onset.

As in other models, this association is not true in earlier decades, although conflicts were commonplace. The consistent findings suggest that across multiple specifications, high horizontal inequalities in the most recent era increase the odds that a country will experience conflict in the next five years.

Part II: Subnational Disparity and Conflict Occurrence in Africa

As we describe above, we examine the effects of subnational inequality in two ways. In Part I above, we looked at the effect of inequality across all regions in a country on that country's likelihood of experiencing violent conflict. In this section, we bring the analysis one level down: our unit of analysis

---

[7] Cederman et al. (2011) propose a lineq2 measure as follows: $lineq2 = \left[\log\left(\overline{d_g}/\overline{D}\right)\right]^2$ We modify this index by replacing the small *d* with the value of the lowest achieving group in the category, and the capital *D* with the value of the highest achieving group in the category. This allows us to bring this measure up and place it at the country level, rather than the level of the subnational unit.

[8] More on measures of inequality is provided in the Technical Annex.

here is the subnational unit rather than the country, so conflict is measured as 1 if it happened within that subnational unit, and 0 if it did not, even if it there was conflict elsewhere in the country. We also modify the predictor variable: instead of measuring inequality between units through a GGini index, we capture inequality as the *difference in mean years of schooling between the subnational unit and the national mean*. This allows us to examine the relationship between a region's mean years of schooling on the likelihood of conflict *in that particular subnational unit*. This method is used to construct the *Subnational Education Inequality and Conflict Dataset*.

Because of the placement of the conflict analysis to the subnational level, our choices are limited in the types of variables that can be used in regression models. We are also cognizant that data availability limits our geographic coverage, as well as our historical time span. Our regression models seek to account for the differences in population and size of subnational unit, as well as their wealth, relative to the national mean. Below, we describe the steps of the data analysis and present our results.

Dataset Construction

As noted above, in the analysis of the effect of subnational disparity and conflict we use the *Subnational Education Inequality and Conflict Dataset*, constructed for a set of 24 African countries. This dataset contains data on mean educational attainment by subnational unit and gender, and conflict. Education inequality data were extracted using the surveys and census datasets, similarly to that described in Part I above. The subnational dataset includes 7,235 data points, drawn from 24 nations in and 237 subnational units in Africa. The unit of analysis in the subnational dataset is the primary administrative region in a nation-state (e.g., province or state). Unlike the national-level dataset, it does not contain data on horizontal inequalities based on ethnic or religious identities. Instead, it focuses on a subnational region's mean years of educational attainment relative to the mean for the nation. The *following steps were taken in the construction of this dataset*:

1. *Mapping.* Subnational units across available household surveys and census datasets compared for all countries for which there were UCDP GED conflict data available. Differences in composition and boundaries were noted.
2. *Alignment*. Subnational units were aligned to match the most recent definition of first-level administrative borders.
3. *Extraction*. Mean years of schooling for males and females extracted for each subnational unit.
4. *Interpolation*. Mean years of schooling interpolated for years between surveys, but only for those subnational units that were vertically aligned. No back projection was performed.
5. *Measurement of inequality*. Absolute distances between the national mean and the mean for the subnational region, in terms of their mean years of schooling, were estimated.

Table 13 shows the country coverage of the subnational dataset. Countries are not represented equally in the dataset, since data for both education inequality were not available on all years for all countries. Countries also have varying numbers of subnational regions. This means that those nations with more subnational units tend to be more represented in the dataset. In particular, as Table 13 shows, our dataset over-represents Nigeria.

Table 13. Countries included in subnational dataset of Education Inequality and Conflict

Country	# Regions	Min Year	Max Year	N	% Dataset
Cameroon	7	2005	2011	168	2.32%
Chad	8	1997	2010	270	3.73%
Democratic Republic of Congo	11	2007	2010	264	3.65%
Republic of Congo	11	2005	2012	245	3.39%
Cote d'Ivoire	11	2006	2012	260	3.59%
Ethiopia	11	2000	2011	264	3.65%
Ghana	10	1993	2011	596	8.24%
Guinea	6	1996	2012	213	2.94%
Kenya	8	1989	2009	456	6.30%
Madagascar	6	1992	2009	216	2.99%
Mali	8	1996	2006	270	3.73%
Morocco	16	1994	2004	384	5.31%
Mozambique	11	1997	2011	386	5.34%
Niger	8	1992	2006	163	2.25%
Nigeria	37	1999	2011	1268	17.53%
Rwanda	10	1992	2010	120	1.66%
Senegal	10	2002	2011	240	3.32%
Sierra Leone	4	2004	2008	96	1.33%
South Africa	9	1996	2007	324	4.48%
Tanzania	6	1992	2005	135	1.87%
Togo	5	1998	2010	165	2.28%
Uganda	5	1991	2011	156	2.16%
Zambia	9	1992	2007	216	2.99%
Zimbabwe	10	1994	2011	360	4.98%

Conflict data. The conflict data for the subnational analysis were extracted from the UCDP Geo-referenced Event Dataset (GED), which covers the African continent spanning 1989-2010 (a much shorter time series than one used for the global analysis). Extensive cleaning and alignment was performed to ensure that each geo-referenced conflict data point could be matched with a subnational unit. This process is documented in the Technical Annex, and more information is available from the authors upon request. We then collapsed all events in the GED dataset at the level of country, year and administrative region to total the number of fatalities and conflict events in each country's subnational region in each year. This converts the dataset to a country-year-region panel dataset that allows it to be merged with the time series dataset that includes data on educational attainment by subnational region over time. Data on total numbers of fatalities in a given calendar year were recoded to denote if any conflict took place, with two alternative codings: a) any battle-related fatality; and, b) over 25 fatalities.

Table 14. Country representation in GED and EIC

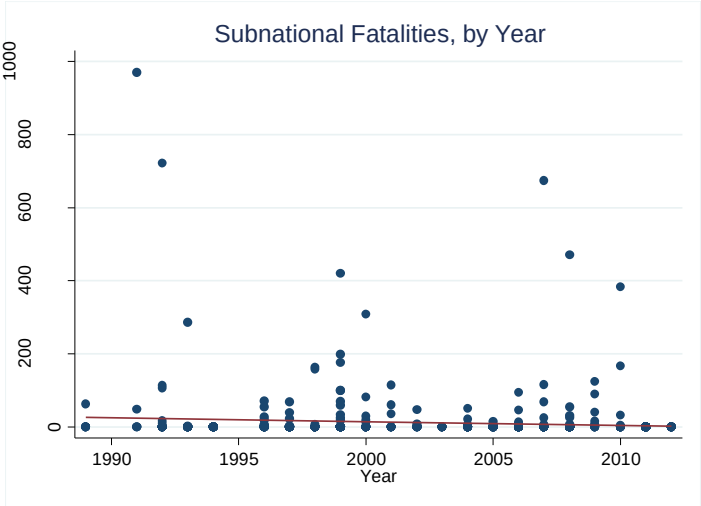
Source	Countries	Subnational Regions	Years	Countries with 1+ Fatalities	Subnational Regions with 1+ Fatalities	Countries with 25+ Fatalities	Subnational Regions with +25 Fatalities
GED	43	515	1989-2010	43	494	38	340
EIC Subnational Dataset	24	237	1989-2012	20	125	11	63

As Table 14. shows, 20 of the 24 countries in the final dataset, representing 125 sub-national regions, have experienced at least one fatality. Of those, 63 sub-national regions in 11 nations, have experienced conflicts resulting in over 25 battle-related fatalities. These numbers are quite a bit smaller than those found in the GED dataset as a whole – while the GED dataset captures data on every battle-related fatality in 43 sub-Saharan nations annually, the SEIC is limited by the availability of educational data. It

only includes data where educational data is available from household survey data and where sub-national regions are consistent over time.

Over the time period, we do not note any major time trends in conflict incidence. There are clearly some years where many more regions experience fatalities, particularly in the late 1990s and late 2000s. We also note some high outliers – with a number of regions experiencing more than 600 deaths in one year (very high outliers are not shown); however, in general, there does not seem to be a time trend, indicating that in the subsequent regression analyses it is not necessary to control for year.

Figure 7. Subnational fatalities by year in final subnational dataset



Descriptive analysis

This section examines the data available for analysis of subnational inequality, including our indicators of subnational inequality and covariates. Less data is available at the level of the subnational regions than is available at the country level; as a result, our covariates are more limited in this analysis.

Subnational Inequality

As noted above, in this section we conceptualize inequality as a direct measure of difference from the national mean. This measure of regional variation in mean years of schooling is summarized in Table 15; it can be interpreted as the difference in mean years of schooling in the region as compared to the national population as a whole. When negative, the value indicates that the country is disadvantaged to the nation as a whole; when positive, it suggests the region is advantaged compared to other regions.

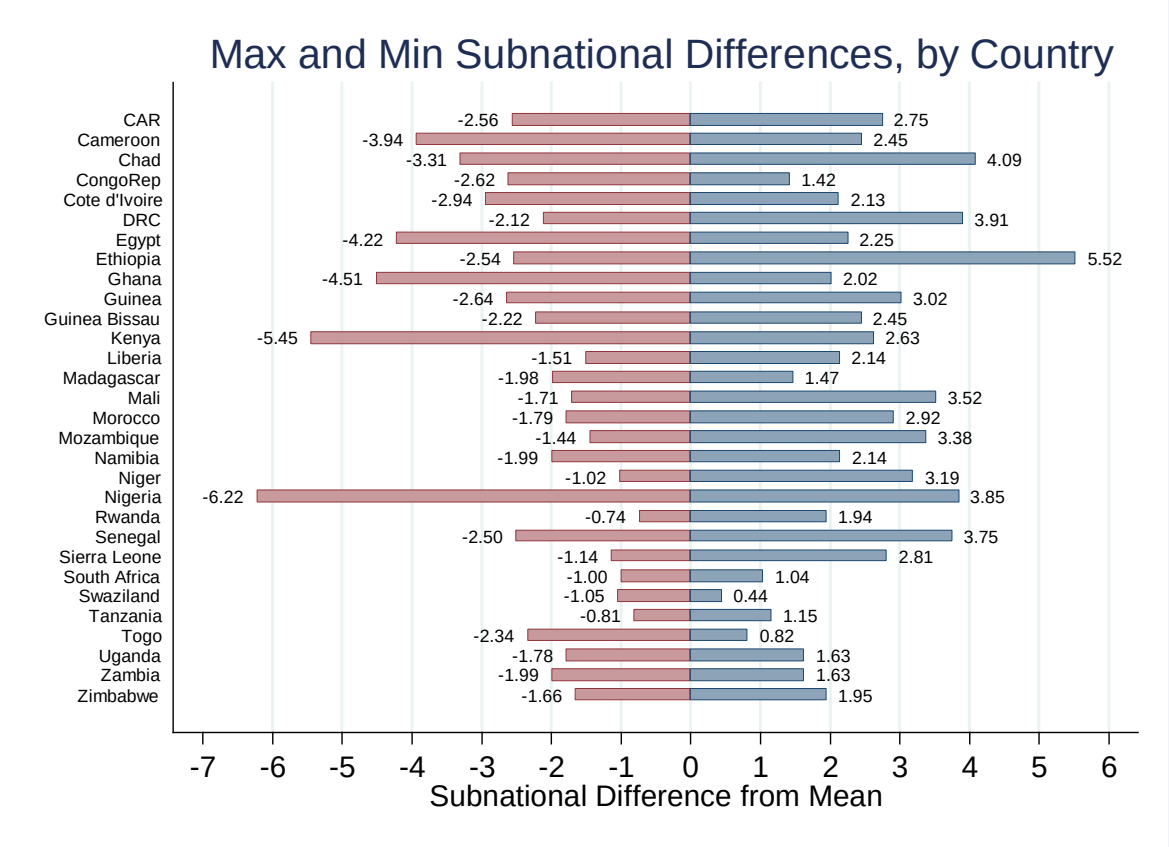
Table 15. Descriptive statistics of subnational gap as absolute difference from the national mean

Gender	Mean	SD	Min	Max	N
Both	-0.18	1.49	-5.99	5.10	690
Female	-0.17	1.58	-5.95	4.84	685
Male	-0.14	1.38	-6.22	5.52	681
Total	-0.17	1.48	-6.22	5.52	2056

Table 15 shows that overall, the mean value is close to zero, although slightly negative. However, the value ranges from a low of six years less than the national mean to a maximum of more than five years above the national mean. Interestingly, the distribution has a smaller range for females than for males.

Figure 8 plots the minimum and maximum subnational difference from the national mean. Countries with the highest inequalities due to privileged regions include Ethiopia (5.52) and Chad (4.09), while countries with severely disadvantaged regions include Nigeria (-6.22) and Kenya (-5.45).

Figure 8. Maximum and minimum values in subnational differences from national mean years of schooling, by country



The measure of subnational disparity is normally distributed, with most observations falling between -3 and +3, or a difference of three years of mean schooling compared to the national mean.

Covariates. We choose from a shorter list of potential covariates for the subnational analysis, as there are fewer predictors of conflict for which there is data that varies at the subnational level annually. We include the following variables:

- *Peace years:* years passed without a single battle-related fatality. This is calculated from the year a country enters the dataset.
- *Population proportion*: the relative size of the subnational unit to the country. This is important if we hypothesize that larger, more densely populated subnational units are more likely to have conflict-related fatalities simply due to their larger populations.

- *Wealth*: unlike GDP per capita, the wealth index we use in this analysis captures the relative wealth of a subnational unit compared to the nation as a whole based on the national distribution of assets. This variable is standardized to a mean of zero and standard deviation of one. The process of creating the wealth index is described in detail in the Technical Annex.[9]

Table 16 provides key descriptive statistics on these covariates. Additional covariates, such as the GDP per capita and the political regime, were not available at levels disaggregated to the subnational unit, and therefore are not included here.

Table 16: Descriptive statistics on covariates in subnational regression models

	Mean	Min	Max	Source
Peace Years	3.80	0	19	GED Dataset
Wealth Index	0.0157	-1.075	2.917	EPDC Calculations
Population percentage	0.1052	0.0013	0.3477	EPDC Calculations

Below we present the process of fitting predictive regression models, as well as the outcomes of regression analysis at the subnational level.

Regression analysis: Subnational disparity

For the analysis of the effects of subnational disparity, we fit multilevel logistic regression models with errors clustered both at the country and subnational unit level. Conflict onset is defined as 1 if at least one battle-related fatality took place in one calendar year in the next five years from the year of inequality observation, and 0 if there were no fatalities.

Our predictor variable also has two specifications. We hypothesize that the magnitude by which the average education in a region differs from the national average matters. Therefore, we create groups of observations, grouping together regions that are one standard deviation from the mean or farther, in years of schooling. This groups the worst off and the best off regions *together*. Using this definition, we measure to what extent *being a substantively higher- or lower- educated region, compared to the rest of the country, contributes to likelihood of conflict*.

However, we know that regions substantively better than the national mean are different from regions substantively worse, and grouping them together misses this important distinction. For this reason, we separate the initial grouping of unequal regions into "bins" that are disadvantaged (1 SD or below from the mean) and advantaged (1 SD or above form the mean). These two indicator variables are entered into regression models with the purpose of capturing the effect of extreme group disadvantage on the likelihood of conflict in that particular region. This allows us to capture potentially differential effects on relative disparity for worse off and better off regions, and examine whether collective deprivation at the subnational level is detrimental to that subnational unit's peace.

[9] We used the existing wealth index for each dataset, where it was available (such as in DHS and MICS surveys), or created our own wealth index using principal components analysis of a range of wealth proxies, different in each country. The universe of household assets and other wealth proxies include: electricity, phone, cell phone, television, kitchen, radio, refrigerator, Internet, email, hot water, computer, washing machine, freezer, VCR, toilet, water source, sewage, trash disposal, land ownership, floor type, automobiles, heat type, air conditioner, and the number of bedrooms.

Table 17. Results of logistic regressions with subnational unit disparity as a predictor of conflict in that subnational unit

	Model 1S	Model 2S	Model 3S	Model 4S	Model 5S	Model 6S
Countries	All	All	W/O Nigeria	All	All	W/O Nigeria
Dependent Variable	1+ Fatality	1+ Fatality	1+ Fatality	25+ Fatalities	25+ Fatalities	25+ Fatalities
Grouping Variables	Country & SN	Country & SN	Country & SN	SN Region	SN Region	SN Region
One SD+ from Mean (0/1)	2.291**			3.569**		
	0.72			1.68		
One SD+ Below Mean (0/1)		2.495*	1.734		4.035*	1.796
		0.93	0.79		2.41	1.42
One SD+ Above Mean (0/1)		1.878	1.596		2.959	2.783
		1.06	0.94		2.15	2.05
Wealth Index	0.455**	0.461**	0.674	0.459*	0.465*	0.684
	0.12	0.12	0.22	0.16	0.17	0.29
Peace Years	0.739***	0.739***	0.770***	0.654***	0.654***	0.667***
	0.02	0.02	0.02	0.03	0.03	0.03
Percent of the Population (%)	0.205	0.229	0.166	0.003	0.003	0.013
	0.67	0.75	0.6	0.01	0.01	0.06
Constant	0.536	0.532	0.375	0.059***	0.058***	0.029***
	0.35	0.35	0.28	0.03	0.03	0.02
Random Effects Parameters						
SD of Subnational Region	2.191***	2.191***	2.427***	4.111***	4.116***	4.297***
	0.45	0.45	0.54	0.39	0.39	0.48
SD of Country Constant						
	2.452***	2.460***	2.907***			
	0.22	0.22	0.31			
N	2590	2590	2127	2590	2590	2127
No. of Subnational Regions	231	231	194	231	231	194
BIC	2116.596	2124.278	1598.965	1318.201	1325.945	941.013

Results

We present two models at the subnational level, distinguished by their definition of subnational disparity: any gap, regardless of higher or lower (Model 1S[10]) and separate measures of disparity, one for disadvantaged and one for advantaged regions (Model 2S). Both of these models are three-level logistic models, with random effects at the country and subnational unit level, and additional variation by year. We retain a fairly large number of observations across the dataset.

Model 1S shows a strong and positive effect of subnational gap (advantage OR disadvantage) on likelihood of conflict across our dataset. A one standard deviation change in average years of schooling, which corresponds roughly to a region with 1.5 mean years of schooling above or below the national

[10] We use the S, for subnational, at the end of the model number to distinguish these models from models fit on a global, country-year panel dataset.

average, is associated with that region having more than twice the odds (2.3) of experiencing conflict than a region with the national mean years of schooling. When we break the regions into two groups, disaggregating disadvantaged and advantaged regions, we see that this is slightly higher for regions worse than one standard deviation away from the mean (an odds ratio of 2.5:1), and it is lower and not statistically significant for regions above the national mean (1.7). The model shows that while both are associated with higher odds of conflict, the odds of conflict are higher in the disadvantaged regions, and it is only in these regions were the odds of conflict are statistically significantly.

To test the robustness of these models, Model 3S excludes Nigeria, which makes up a disproportionate percent of the observations in our dataset, and also happens to be a country with the most disadvantaged regions and a conflict-affected nation. Model 3S shows that after excluding Nigeria, the sign of the coefficients remains positive – meaning both advantaged and disadvantaged regions are expected to experience higher odds of conflict, but the coefficient on disadvantaged regions drops significantly and is no longer significant. This finding suggests that the relationship between subnational inequality in educational attainment and the likelihood of that region experiencing battle-related fatalities is not robust. Given this finding, our results are inconclusive on this question.

In the next set of models (Model 4S-6S), we examine the effect of subnational inequality on violent conflict with a higher threshold for conflict, now defined as 25 battle-related fatalities or higher. The 25-plus battle death minimum is the same dependent variable used in the cross-national models above, allowing us to examine whether the trends we noted at the national level are comparable at the subnational level as well. Because of the lack of variation in the dependent variable, we fit a random effects model for only the subnational unit, not a two-level model as in Models 1S-3S. To assure that these models are comparable to Models 1S-3S, we tested both model fit and the variable coefficients, and find that while random intercepts at both the country and subnational unit do improve model fit slightly, it has little effect on the coefficients or substantive conclusions. Therefore, we are confident that Models 4S-6S are comparable to prior models, and that the only major difference is the dependent variable.

In Models 4S and 5S, we find that subnational regions that are both advantaged and disadvantaged relative to the national mean have higher odds of experiencing conflicts, and that the odds are higher in disadvantaged regions. However, Model 6S suggests that, as in the first set of models, removing Nigeria from the models makes the likelihood of conflict not statistically significantly different from zero.

Combined, the subnational analyses suggest that regions with lower mean education levels are no more or less likely to experience battle-related fatalities – either low or high levels of fatalities – than regions with mean levels of education.

There are a number of reasons why we do not find a statistically significant effect – the most obvious explanation is simply that the outbreak of conflict at the subnational level is quite different than at the national level. For example, fatalities may not fall neatly into subnational borders. Additionally, although we try to control for population, urban areas tend to be both more advantaged and more populous, meaning they could bear the brunt of battle-related deaths, even if conflict broke out first in other regions. Our subnational data is also limited somewhat by data availability, which could affect the findings– a country's administrative regions, on which our dataset is based, may be somewhat arbitrary – they may be divided into many more or fewer units than are understood to be politically relevant in the national context.

Discussion

Our findings show that there is a robust and consistent statistical relationship between higher levels of inequality in educational attainment between ethnic and religious groups and the likelihood that a country will experience violent conflict at the global level. Our interaction effects with time show that this relationship is large and robust in the most recent decade, while in earlier decades there seems to be little statistically significant relationship between educational inequality and the likelihood of conflict. It is important to emphasize that our findings are not necessarily pointing to a direct and causal relationship, i.e. that education inequality between groups is the cause of violent conflict. Further, education inequality may serve as a proxy of inequality in access to other services or political and economic privileges (beyond the basic income Gini for which we control). However, to the extent that a strong theoretical linkage can be made between educational inequality, on the one hand, and economic and political disempowerment on the other, one can argue that there may be an indirect yet causal relationship whereby systematic inequality in education experienced by some subgroups and the formation of group-based grievances eventually lead to conflict.

As our literature review (FHI 360/ EPDC 2015) describes, there are a number of potential avenues through which education inequality may directly or indirectly lead to conflict. The link between education and future economic productivity and wellbeing is perhaps the most often cited; however, a number of authors have pointed to the crucial role that education plays in the formation of social cohesion and national identity. Educational inequality, in this argument, leads to imbalances in the societal fabric and reinforces the regression to group allegiances. Education is an inherently political process, and hence inequality in education is necessarily linked to political disempowerment and disadvantage in other spheres.

It is also noteworthy that the relationship between ethnic and religious educational inequality and violent conflict, as it is observed in our study, has changed over time. Although we cannot be certain why this interaction effect is present, it is possible that the social consequences and meanings associated with inequality have changed over time, such that the same magnitude in our inequality measure denotes much deeper levels of exclusion. In the early decades of our time series, horizontal inequalities in education were objectively much higher than they were in the 2000s. However, high levels of education inequality may not have been considered a sufficient reason for grievance in the 1970s and 1980s, when inter-group inequality was commonplace and access to education not construed as a universal right. Over the past two decades, important changes have taken place in countries around the world– mass schooling has expanded, access to education has been accepted as a basic right for all children, and higher levels of schooling have become increasingly important for entry into the labor market. Consequently, high levels of inter-group inequality in educational attainment may signal greater levels of disempowerment and systematic exclusion of some groups from future economic opportunities. It may also be perceived as one way that the nation-state is failing to meet its basic responsibilities to provide social services. All of these factors mean that one ethnic or religious group could perceive educational inequality as an injustice, or a reason for discontent.

Additionally, we find that the relationship between subnational educational inequalities and the likelihood of conflict is present and stable across all five decades. This relationship is statistically significant after controlling for important covariates even on a much smaller dataset. This finding suggests two possible interpretations: 1) either subnational inequalities are somehow different than ethnic or

religious horizontal inequalities; or 2) although the relationship between horizontal inequality and conflict is likely the same for all group types, due to issues of country representation and data availability, the relationship in early decades only appears in the subnational analysis. At this point, our data cannot distinguish between these two possible explanations. It is possible that our dataset on subnational inequalities, which is much smaller and does not have back projections, is not representative of all nations over time, and that greater country coverage and longitudinal data would find results more in line with those with ethnic and religious inequalities. We cannot say definitively without more data on subnational inequalities.

Nonetheless, it is also possible that subnational inequalities – based on administrative regions – operate differently than those based on religious or ethnic group identities. In some countries, where ethnic groups live in particular regional pockets, the two may be synonymous with one another. However, it is also possible that ethnic and religious groups lived in various parts of a country and identity lines do not fit neatly into subnational borders. Instead, in these countries, individuals may have shared grievances over underinvestment by the state in their subnational region – perhaps isolated by geography or marginalized due to long distances from the capital. In these regions, conflict may arise out of territorial disputes or discontent at the state's lack of investment in social services in their region.

Recommendations: a research agenda

This study points to a number of areas for future research. First is the need for greater data availability. Our analysis, while drawing on the most comprehensive dataset on horizontal inequalities available to-date, lacks sufficient representation from a number of regions, particularly Western and developed nations and nations in the Middle East and North Africa. Additionally, at the subnational level with the SEIC dataset, we were only able to find longitudinal educational data on 24 of the 43 countries covered by the GED. More data is needed to understand if our findings are affected by data coverage.

In addition, future research is needed to replicate findings using different definitions and specifications of key variables for both conflict and educational inequality. First, it is important to remember that in the global analysis (Part I), the dependent variable is defined somewhat narrowly as the onset of intrastate conflict in which the state is one party to the conflict. As such, it fails to capture other types of civil conflict that may be fueled by intergroup inequality or grievances among various non-state actors. Hence, future studies must replicate results using other definitions of conflict in order to investigate whether the relationship between horizontal inequality and the likelihood of conflict differs by the type of civil conflict.

Secondly, our dataset measures educational inequality using a single measure: educational attainment. While it is a strong measure of the stock of human capital accumulated in a group and has been used extensively in the literature, attainment captures only one side of potential educational inequality. It leaves out other important dimensions, including levels of resources and the quality of educational inputs, which have implications for future economic productivity and civic participation. More research on the relationship between these aspects of education and violent conflict is also needed.

In addition, the time trend found in our analysis also deserves greater attention. More research is needed to understand why educational inequality in the most recent decade is associated with conflict when this does not seem to be the case in the preceding historical period. In particular, further research could

investigate whether the differential effect seems to lie primarily in the changing nature and significance of horizontal inequality and if educational inequalities are increasingly linked to grievances.

Finally, an important area for future research is the reverse relationship – the effect of conflict on educational inequality. We know that inequality and conflict may operate as a cyclical relationship, with educational inequality exacerbating discontent and conflict having a disproportionately negative impact on some regions and populations than others. In the future, researchers must examine the relationship between the experience and duration of conflict in a given country and horizontal educational inequalities.

References

Bartusevicius, H. (2014). The inequality-conflict nexus re-examined: Income, education and popular rebellions. *Journal of Peace Research*, *51*(1), 35-50.

Brown, G. K. (2009). Regional autonomy, spatial disparity, and ethnoregional protest in contemporary democracies: A panel data analysis, 1985-2003. *Ethnopolitics*, *8*(1), 47-66.

Cederman, L., Weidmann, N. B., & Gleditsch, K. (2011). Horizontal inequalities and ethnonationalist civil war: A global comparison. *American Political Science Review, 105*(3), 478-495.

Collier, P., & Hoeffler, A. (2004). Greed and grievance in civil war. *Oxford Economic Papers, 56(4),* 563-595.

Fearon, J. D., & Laitin, D. D. (2003). Ethnicity, insurgency, and civil war. *American Political Science Review, 97*(1), 75-90.

Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer (2013), "The Next Generation of the Penn World Table."

FHI 360 Education Policy and Data Center. (2015). *Horizontal Education Inequality and Violent Conflict: a Literature Review*. New York, NY: UNICEF.

Hegre, H., & Sambanis, N. (2006). Sensitivity analysis of empirical results on civil war onset. *Journal of Conflict Resolution, 50*(4), 508-535.

Mancini, L., Stewart, F., & Brown, G. K. (2008). Approaches to the measurement of horizontal inequalities. In F. Steward (Ed.), *Horizontal inequalities and conflict: understanding group violence in multiethnic societies (pp.85-105)*. Palgrave Macmillan.

Montalvo, J. G., & Reynal-Querol, M. (2005). Ethnic polarization, potential conflict, and civil wars. *American Economic Review*, *95*(3), 796-816.

Østby, G. (2008). Polarization, horizontal inequalities and violent civil conflict. *Journal of Peace Research, 45*(2), 143-162.

Polity IV. (2012). "Polity IV Project: Political Regime Characteristics and Transitions, 1800-2012".

Ross, M. L. (2005). Resources and rebellion in Aceh, Indonesia. In P. Collier & N. Sambanis (Eds.), *Understanding civil war: Evidence and analysis, Vol. 2. Europe, Central Asia, and other regions* (pp. 35-58). Washington, DC: The World Bank.

Stewart, F., Brown, G., & Mancini, L. (2010). *Monitoring and measuring horizontal inequalities*. Oxford: Centre for Research on Inequality, Human Security, and Ethnicity.

Themnér, L. & Wallensteen, P. (2013). Armed Conflict, 1946-2012 [UCDP Armed Conflict data resource]. *Journal of Peace Research, 50*(4), 509-521.

United Nations Population Division. Population and Development Database (2012). New York, N.Y.: UNPD

UNU-WIDER. (2014). "World Income Inequality Database (WIID3.0b)." Helsinki, Finland: United Nations University.

Vreeland, J.R. (2008) The effect of political regime on civil war: Unpacking anocracy. *Journal of Conflict Resolution, June 2008 vol 52 no 3,* pp. 401-425.

The World Bank. (2014). World Development Indicators. Washington, D.C.: The World Bank.

Appendices

Appendix A. Data availability: Global Dataset of Education Inequality and Conflict

Table 18. Geographic coverage in the Education Inequality and Conflict Dataset

Country	Ethnic	Religious	Sub-National	Total Years	Start Year	End Year	Ever Conflict
Afghanistan	1	0	0	38	1971	2008	1
Albania	0	1	1	41	1968	2008	0
Argentina	0	0	1	4	1970	2001	1
Armenia	0	0	1	1	2000	2000	0
Austria	0	1	0	32	1961	2001	0
Azerbaijan	0	0	1	1	2006	2006	0
Bangladesh	0	1	1	32	1961	2001	1
Belarus	1	0	0	31	1969	1999	0
Benin	1	1	1	43	1966	2008	0
Bolivia	1	0	1	41	1961	2001	1
Brazil	1	1	1	33	1960	2000	0
Burkina Faso	1	1	1	46	1963	2008	1
Burundi	0	1	1	44	1965	2008	1
Cambodia	0	0	1	1	1998	1998	0
Cameroon	1	1	1	48	1961	2008	1
Canada	1	1	0	32	1961	2001	0
Central African Republic	1	1	1	42	1965	2006	1
Chad	1	1	1	45	1964	2008	1
Chile	0	1	1	42	1960	2002	1
Colombia	1	0	1	46	1963	2008	1
Congo DR	1	1	1	42	1967	2008	1
Congo Rep	1	1	1	44	1965	2008	1
Costa Rica	0	0	1	4	1963	2000	0
Cote d'Ivoire	1	1	1	43	1966	2008	1
Dominican Republic	0	1	1	41	1962	2002	1
Ecuador	1	0	1	48	1961	2008	0
Egypt	0	1	1	42	1966	2008	1
El Salvador	0	0	1	2	1992	2007	0
Ethiopia	1	1	1	49	1960	2008	1
Fiji	1	1	1	42	1966	2007	0
Gabon	1	1	1	49	1960	2008	1
Gambia	1	0	1	41	1966	2006	1
Germany	0	1	0	28	1960	1987	0
Ghana	1	1	1	46	1963	2008	1
Guatemala	1	0	1	21	1967	1987	0
Guinea	1	1	1	44	1965	2008	1
Guinea-Bissau	1	1	1	41	1966	2006	1
Guyana	1	1	0	40	1969	2008	0
Haiti	0	1	1	48	1961	2008	1
Honduras	1	1	0	37	1972	2008	0
India	0	1	1	37	1963	1999	1
Indonesia	1	1	1	49	1960	2008	1
Israel	0	1	1	34	1962	1995	1
Jamaica	1	1	1	41	1961	2001	0
Kazakhstan	1	1	1	44	1965	2008	0
Kenya	1	1	1	46	1963	2008	1
Kyrgyz Republic	1	1	1	41	1966	2006	0
Laos	1	1	1	43	1966	2008	1
Lesotho	0	1	0	40	1969	2008	1
Liberia	1	1	1	42	1966	2007	1
Macedonia	1	1	1	44	1965	2008	1

Country	Ethnic	Religious	Sub-National	Total Years	Start Year	End Year	Ever Conflict
Madagascar	0	1	1	47	1962	2008	1
Malawi	1	1	1	41	1960	2000	0
Malaysia	1	1	1	41	1960	2000	1
Mali	1	0	1	33	1966	2006	1
Mexico	1	0	1	42	1960	2005	1
Moldova	1	0	1	41	1965	2005	1
Mongolia	1	1	1	44	1965	2008	0
Morocco	0	0	1	3	1982	2004	0
Mozambique	1	1	1	46	1963	2008	1
Namibia	1	1	1	46	1962	2007	0
Nepal	1	1	1	43	1966	2008	1
Nicaragua	0	1	1	32	1965	2005	1
Niger	1	0	1	45	1962	2006	1
Nigeria	1	1	1	41	1968	2008	1
Pakistan	1	0	0	43	1966	2008	1
Panama	1	0	1	32	1960	2000	1
Peru	1	1	1	47	1962	2008	1
Philippines	1	1	1	49	1960	2008	1
Portugal	0	1	0	32	1961	2001	0
Romania	1	1	0	41	1962	2002	1
Rwanda	1	1	1	47	1962	2008	1
Sao Tome Principe	0	1	0	30	1979	2008	0
Senegal	1	1	1	47	1962	2008	1
Serbia	1	1	1	43	1966	2008	1
Sierra Leone	1	1	1	45	1964	2008	1
South Africa	1	1	1	42	1966	2007	1
Sri Lanka	1	1	1	21	1967	1987	1
Suriname	1	1	1	41	1966	2006	1
Swaziland	0	1	1	41	1966	2006	0
Switzerland	0	1	0	32	1960	2000	0
Tanzania	0	1	1	44	1962	2005	0
Thailand	0	0	1	4	1970	2000	1
Togo	1	1	1	43	1966	2008	1
Trinidad and Tobago	1	1	1	41	1966	2006	1
Turkey	0	1	0	31	1968	1998	1
Uganda	1	1	1	48	1961	2008	1
Ukraine	0	1	1	41	1967	2007	0
United States	1	0	1	14	1960	2000	1
Uruguay	1	1	1	42	1963	2006	1
Uzbekistan	1	0	1	21	1976	1996	1
Venezuela	0	0	1	4	1971	2001	1
Vietnam	1	1	1	43	1966	2008	0
Zambia	1	1	1	46	1962	2007	0
Zimbabwe	0	1	1	45	1964	2008	1

Appendix B: Sensitivity Checks

Table 19. Regression results on final model (Model 4), with an alternative specification of decade bins

	Model 15
Model Type	Random Effects
Group Identity	Ethnic or Religious
GGINI: Horizontal Inequality (1996-2008)	2.472**
	0.69
1985-1995	2.089***
	0.43
1975-1984	1.466
	0.4
1960-1974	0.908
	0.34
1985-1995 # Group GINI	0.268***
	0.07
1975-1984 # Group GINI	0.462**
	0.13
1960-1975 # Group GINI	0.769
	0.21
GDP per capita (logged)	0.726+
	0.13
Peace Years	0.910***
	0.02
Peace Years Squared	1.003***
	0.0
Population (logged)	1.527*
	0.31
Democracy (0/1)	1.042
	0.25
Anocracy (0/1)	2.045***
	0.38
Ethnic Groups	0.913
	0.11
Constant	0.089***
	0.06
Random Effects Parameters	
S.D. of Constant	2.499***
S.E.	0.38
N	2483
N. Countries	79
BIC	1937.2

Table 20. Regression results: final model with robustness checks on length of time series and model specification

	Model 16	Model 17	Model 18
Robustness Check	No Long Projections	Five Decades Only	Fixed Effects
Model	Random Effects	Random Effects	Fixed Effects
Group Identity	Ethnic or Religious	Ethnic or Religious	Ethnic or Religious
Gender	Both	Both	Both
GGINI: Horizontal Inequality	2.074+	3.079***	2.435**
	0.87	0.96	0.80
1990s	1.786*	1.522+	2.402***
	0.44	0.33	0.63
1980s	4.164***	2.566**	6.595***
	1.5	0.74	2.69
1970s	3.449*	2.019+	9.062***
	1.87	0.76	5.11
1960s	12.738*	1.55	14.758***
	13.9	0.74	10.73
1990s # Group GINI	0.310***	0.322***	0.331***
	0.1	0.09	0.1
1980s # Group GINI	0.149***	0.229***	0.304***
	0.06	0.07	0.1
1970s # Group GINI	0.271*	0.377**	0.545+
	0.16	0.12	0.17
1960s # Group GINI	9.741	0.752	0.966
	13.54	0.25	0.33
GDP per capita (logged)	0.938	0.771	1.132
	0.25	0.14	0.24
Peace Years	0.879***	0.901***	0.851***
	0.02	0.02	0.02
Peace Years Squared	1.004***	1.003***	1.005***
	0	0	0
Population (logged)	1.974*	1.869*	8.918**
	0.55	0.46	6.26
Democracy (0/1)	1.172	1.411	1.547+
	0.38	0.35	0.39
Anocracy (0/1)	2.111**	2.439***	2.473***
	0.54	0.46	0.47
Ethnic Groups	1.354	1.027	
	0.26	0.13	
Constant	0.006***	0.035***	
	0.01	0.03	
Random Effects Parameters			
S.D. of Constant	3.119***	2.543***	
S.E.	0.52	0.42	
N	1782	2515	1958
N. Countries	78	71	55
BIC	1358.842	1972.787	1538.346

Table 21. Correlations between key variables in the global regression models (see Part I).

	Conflict Onset	Group GINI	Log GDP	Log Pop	Log Pop Dens.	Youth Pop	Democracy	Anocracy	Ed GINI	Wealth GINI	Ethnic Frac.	% Mnt. Terrain	Oil producer	Ed Spend.
Predictor Variable														
Group GINI	-0.006	1.000												
Covariates														
Log GDP	-0.138	-0.369	1.000											
Log Population	0.274	-0.033	-0.105	1.000										
Log Pop. Density	0.077	-0.126	-0.133	0.366	1.000									
Youth Pop (%)	-0.002	-0.053	-0.178	0.017	0.083	1.000								
Democracy	-0.047	-0.275	0.429	0.036	0.189	0.051	1.000							
Anocracy	0.104	0.024	-0.160	0.082	-0.012	0.155	-0.400	1.000						
Education GINI	0.108	0.458	-0.558	-0.057	-0.043	-0.166	-0.364	0.029	1.000					
Wealth GINI	-0.152	0.061	0.046	-0.217	-0.181	0.162	-0.036	-0.058	0.037	1.000				
Ethnic Fraction.	0.148	0.298	-0.258	0.057	-0.300	0.044	-0.266	0.168	0.204	-0.016	1.000			
Mountain Terrain	0.078	-0.203	0.040	0.323	0.093	0.098	0.072	-0.009	-0.137	-0.055	-0.266	1.000		
Oil Production	0.027	-0.186	0.536	0.207	-0.061	-0.011	0.147	-0.022	-0.353	-0.072	-0.017	-0.016	1.000	
Education Spending (% GDP)	-0.096	-0.043	0.079	0.025	0.036	-0.115	0.160	-0.139	-0.040	0.018	-0.008	0.076	-0.158	1.000
Mean Ed. Attainment	-0.108	-0.565	0.690	0.062	0.099	0.046	0.461	-0.087	-0.818	-0.072	-0.211	0.164	0.423	-0.009

Subnational Analysis: Regional Variation by Country

Table 22. Overview of regional variation from country mean

Country Name	Mean	Min	Max	SD	N
Cameroon	0.43	1.74	-3.94	2.45	42
Central African Republic	-0.61	1.16	-2.56	2.75	48
Chad	-0.23	1.87	-3.31	4.09	72
Democratic Republic of Congo	-0.15	1.26	-2.12	3.91	66
Republic of Congo	-0.99	0.98	-2.62	1.42	64
Cote d'Ivoire	-0.56	1.26	-2.94	2.13	66
Egypt	0.06	1.21	-4.22	2.25	79
Ethiopia	0.71	1.75	-2.54	5.52	66
Ghana	-0.51	1.61	-4.51	2.02	149
Guinea	-0.21	1.37	-2.64	3.02	54
Guinea-Bissau	-0.65	1.27	-2.22	2.45	27
Kenya	-0.17	1.65	-5.45	2.63	114
Liberia	-0.24	1.35	-1.51	2.14	12
Madagascar	-0.15	0.80	-1.98	1.47	54
Mali	0.02	1.10	-1.71	3.52	72
Morocco	0.02	1.08	-1.79	2.92	96
Mozambique	0.11	1.13	-1.44	3.38	99
Namibia	-0.22	1.11	-1.99	2.14	39
Niger	0.25	1.21	-1.02	3.19	47
Nigeria	-0.22	2.21	-6.22	3.85	319
Rwanda	0.10	0.66	-0.74	1.94	30
Senegal	-0.22	1.60	-2.50	3.75	60
Sierra Leone	0.04	1.50	-1.14	2.81	24
South Africa	-0.09	0.50	-1.00	1.04	81
Sudan	-0.08	0.58	-1.05	0.44	12
Swaziland	0.03	0.55	-0.81	1.15	36
Tanzania	-0.42	0.90	-2.34	0.82	45
Togo	-0.03	0.93	-1.78	1.63	39
Uganda	-0.35	0.99	-1.99	1.63	54
Zambia	-0.12	0.85	-1.66	1.95	90
Zimbabwe	-0.17	1.48	-6.22	5.52	2056