LICOS Centre for Institutions and Economic Performance

Centre of Excellence

LICOS Discussion Paper Series

Discussion Paper 275/2011

# Measure for Measure:
# How Well Do We Measure Micro-level Conflict Intensity?

*Marijke Verpoorten*

**Katholieke Universiteit Leuven**

LICOS Centre for Institutions and Economic Performance
Waaistraat 6 – mailbox 3511
3000 Leuven

BELGIUM

TEL:+32-(0)16 32 65 98
FAX:+32-(0)16 32 65 99
http://www.econ.kuleuven.be/licos

# Measure for Measure:

# How Well Do We Measure Micro-level Conflict Intensity?

Marijke Verpoorten*

University of Leuven

January, 2011

**Abstract**

Rich measures of micro-level violent conflict intensity are key for successfully providing insight into the legacy of civil war. Yet, the debate on how exactly conflict intensity should be measured has just started. This paper aims to fuel this awakening debate. It is demonstrated how existing and widely available data - population census data - can provide the basis for a useful measure of micro-level conflict intensity, i.e. a fine Wartime Excess Mortality Index (WEMI). In contrast to measures that are based on news reports or data from transitional justice records, WEMI is relatively neutral to the cause of excess mortality, giving equal weight to victims belonging to the conquering and defeated party, to victims of large-scale massacres and dispersed killings, to victims in easily accessible locations and remote areas, and to direct and indirect victims of violence. The measure is illustrated for the case of Rwanda and it is shown that in a straightforward empirical application of the impact of armed conflict on schooling different measures for micro-level conflict intensity yield strikingly different results.

JEL: C81, 015, C21

Armed Conflict, Micro-level conflict intensity measures, Difference-in-Difference, Rwanda, Schooling

*Research scholar of the fund for Scientific Research - Flanders, Belgium (FWO), Center for Economic Studies, Centre for Institutions and Economic Performance LICOS, KULeuven. e-mail: marijke.verpoorten@econ.kuleuven.be

## 1 Introduction

The challenge ahead for the micro-empirical literature on the legacy of civil war lies in providing in depth insight into the various underlying mechanisms of the alleged conflict trap. As argued in a recent overview by Blattman and Miguel (2010): "The leading question is not whether wars harm human capital stocks, but rather in what ways, how much, for whom, and how persistently", or "The evidence is weakest where it is arguably most important: in understanding civil wars' effects on institutions, technology and social norms." Blattman and Miguel (2010) further argue that new data is key to successfully embarking on these challenges: "A major goal of civil war researchers should be the collection of new data". Or more specifically, in the words of Restrepo, Spagat, and Vargas (2006): "Conflict researchers should prioritize the construction of more micro datasets that will facilitate detailed studies of conflict intensity and its dynamics".

Taking this advice to heart, this paper demonstrates how existing and widely available data - population census data - can form the basis of a useful measure of micro-level conflict intensity.

Rwanda is taken as an illustration. This country in Africa's turbulent Great Lakes region experienced several forms of violence in the nineties, including genocide, civil war, reprisal killings and (counter)insurgency (i.e. rural guerilla warfare). In order to measure Rwanda's complex and multidimensional conflict cycle, I develop a spatial index of wartime excess mortality relying on the 1991 and 2002 Rwandan population census. In particular, I subject a number of community level wartime excess mortality proxies (1991-2002 differences in mortality of sons and daughters, widowhood and orphanhood, and 2002 disability due to armed conflict) to principal component analysis (PCA). The first principal component (PC) provides us with a Wartime Excess Mortality Index (WEMI) on a less to more scale for 145 administrative units ("communes")[1].

The usefulness of the WEMI as a conflict intensity measure in micro-empirical applications is threefold. First, given the complete coverage of the population in the census, the WEMI yields a very fine measure, i.e. at the level of small administrative units, which allows to capture within province variation in the intensity of armed conflict. Second, given the wide availability and uniformity of population census data across even the least developed

[1] In the administrative subdivision of Rwanda Anno 1994, "prefectures" are followed by "communes", "sectors" and "cells". In a series of subsequent reforms in 1996 and 2002, prefectures and communes were replaced by respectively provinces and districts. Sectors are, after the cell, the lowest administrative units in Rwanda with an average population size of about 5000 inhabitants. A commune has an average size of 134 squared km and counts on average 5.8 sectors.

countries, the proposed measure can be generalized to other countries that have experienced armed conflict. A uniform conflict intensity measure is valuable in the sense that it enhances the comparability of different country case studies and therefore the general lessons that can be drawn from these studies. A third useful characteristic of the WEMI is its neutrality, i.e. it is relatively neutral towards the cause of excess mortality. In contrast to conflict intensity measures derived from transitional justice records or news reports, the WEMI gives equal weight to victims belonging to the conquering and defeated party, to victims of large-scale massacres and dispersed killings, to victims in easily accessible locations and remote areas, and to direct and indirect victims of violence[2].

While this neutrality is useful in itself, e.g. to reduce omitted variable bias when estimating the impact of conflict, it can also be the basis for the construction of a multi-dimensional conflict intensity measure. For the case of Rwanda, I develop a two-dimensional index by augmenting the "neutral" set of excess mortality proxies from the population census data with genocide proxies derived from the records of the gacaca (the Rwandan transitional justice system). Subjecting both sets of variables jointly to PCA results in a first and second PC with a clear interpretation that allows to distinguish between, on the one hand, the 1994 genocide excess mortality (GEMI) and, on the other hand, other sources of wartime excess mortality, including the 1991-1994 civil war, the 1994-1998 (counter)insurgency and the 1994-1998 refugee crisis (CEMI).

There are always two sides to a coin and WEMI comes with at least two drawbacks. First, WEMI may suffer from survival bias since it is based on information inferred from the surviving population, more precisely from close relatives of those who died. Hence, WEMI may be biased downward for communities were many families were entirely exterminated. In the case of Rwanda survival bias is likely to be present and I will discuss how this bias can be attenuated using information on the location of mass graves. A second drawback is that other events unrelated to conflict may explain wartime excess mortality (e.g. a local harvest failure, a region-specific mortality trend). A way to account for this is to revert to Instrumental Variable Estimation (IVE) when using the measure in an econometric application, an approach that has by now become standard in micro-empirical studies on the legacy of violent conflict.

The usefulness of the proposed one- and two-dimensional measures of conflict intensity

[2] It is estimated that the number of indirect deaths of conflict is six times larger than the number of battle-related direct deaths. For a discussion of the causes and extent of the indirect deaths from war, see the 2009/2010 Human Security Report (part II) (Human Security Report Project (2010)).

is illustrated with an empirical application of the impact of armed conflict on schooling attainment in Rwanda. The identification of the impact relies on a difference-in-difference-in-difference estimation of years of education of a young (6-22) and an older (23-50) age cohort in the 1991 and 2002 population census, with the treated group being the young age cohort in the 2000 census residing in high conflict intensity regions. To account for the possible endogeneity of conflict intensity, I use an instrumental variable approach. It is shown that the results obtained with the newly proposed one- and two-dimensional measures deviate strongly from the results obtained using a large number of alternative measures of conflict intensity in Rwanda that are either less fine or less neutral or both.

This application is closely related to Akresh and de Walque (2008) who study years of education of a young and an older age cohort in the 1992 and 2000 DHS survey. The authors find that children exposed to armed conflict completed close to one-half year less education which corresponds to a 18.3% drop relative to the average educational achievement. I don't contest this result. In fact, I find a qualitatively similar result using two rounds of population census data. However, lacking detailed data on different forms of violence in Rwanda, the authors implicitly attribute the entire estimated schooling deficit to genocide. Using the two-dimensional measure (GEMI, CEMI), I find that at least half of the estimated schooling deficit can be attributed to other events in the Rwandan conflict cycle.

The remainder of this paper is structured as follows. Section 2 gives an overview of the conflict intensity measures used in the micro-empirical literature on the legacy of violent conflict. Three types of measures are defined. Each type is evaluated on its strengths and weaknesses. Section 3 lays out the method for the newly proposed measure. Section 4 illustrates the method for the Rwandan conflict cycle. Section 5 provides an application studying the impact of violent conflict on schooling in Rwanda. Section 6 compares the results with previous measures. Section 7 concludes.

## 2 The micro-empirical literature on the legacy of violent conflict: a typology according to the conflict measure used

Below, I give an overview of micro level studies that analyze the impact of civil war on socioeconomic outcomes[3]. For reasons of parsimony, studies that focus on two particular groups of individuals, i.e. refugees and ex-combatants, are excluded from the overview.

[3] Civil war is defined as a war between organized groups within a single nation state having more than 1,000 battle deaths in a single year (Gleditsch, Wallensteen, Eriksson, Sollenberg, and Strand (2002)).

Among the studies discussed, three main types are defined according to the measurement of conflict exposure: conflict exposure in time, conflict exposure in space, and household conflict experience. Table 1 summarizes the studies by type.

---

Insert Table 1 about here

---

**Type I: conflict exposure in time** The first type of study measures conflict exposure by combining information on the conflict's timing with birth dates of the surveyed population, thus identifying the affected age cohort. This method is often feasible because in most cases the conflict's start and end date are known by reasonable approximation. In addition, commonly executed nationwide surveys include birth dates of the sampled population. The most convincing results are obtained when scholars can use difference-in-difference estimates relying on two nationwide surveys, one prior and one after the conflict occurred, with the treated group being the affected age cohort in the post-conflict survey. However, even in those cases, Type-I studies face two drawbacks. First, the outcomes studied are restricted to age-related individual characteristics, e.g. schooling attainment, height-for-age z-scores and fertility. Second, one cannot exclude that the results are driven by another event in the same time span or by a time trend. To control for the latter, type-I studies often use a difference-in-difference-in-difference estimation, relying on variation in conflict intensity across time as well as space.

Examples of type-I studies are Akresh and de Walque (2008), Akresh, Verwimp, and Bundervoet (forthcoming), Alderman, Hoddinott, and Kinsey (2006), Bundervoet, Verwimp, and Akresh (2009), Chamarbagwala and Morán (2008), de Walque (2004) and Leon (2010). Among these seven studies, four focus on schooling outcomes, while three look at height-for-age z-scores (see Table 1 for details). Leon (2010) uses the finest measure of conflict intensity. He estimates the short and long term effects of the Peruvian civil strife on educational achievement using both the conflict's variation in time and space. The spatial variation is based on the number of human rights violations across 1833 districts provided by the Peruvian Truth and Reconciliation Commission. de Walque (2004) studies the impact of the Cambodian Khmer Rouge terror regime on population structure, health status as well as schooling levels. In contrast to the other studies, this one does not combine variation of conflict intensity in time with its variation in space. All other studies do so and therefore overlap with studies of type II.

**Type II: conflict exposure in space** A first set of studies using spatial variation in conflict intensity relies on event data to construct a dummy variable taking one for provinces heavily affected by violent conflict and zero otherwise. The event data mostly stem from journalists' or human rights organizations' reports. It is up to the researcher to interpret the reports, which introduces a degree of arbitrariness. In addition, event data do not systematically cover all areas of a country and are therefore inherently biased. For example, areas that are relatively well accessible or areas in which large scale massacres took place may receive more and better coverage compared to less accessible areas and areas where killings were dispersed. Examples of studies that define a conflict dummy based on event data are two previously mentioned type-I studies (Akresh, Verwimp, and Bundervoet (forthcoming) and Bundervoet, Verwimp, and Akresh (2009)) as well as Bundervoet (2007) and Justino and Verwimp (2006).

A second set of type-II studies have constructed richer measures of geographic conflict intensity based on very diverse sources. This set includes three previously mentioned type-I studies (Leon (2010), Chamarbagwala and Morán (2008) and Akresh and de Walque (2008)), as well as an additional three studies: González and Lopez (2007), Miguel and Roland (Forthcoming) and Li (2007). González and Lopez (2007) study the impact of violence in Colombia on farm productivity. They make use of principal component analysis to construct an index of political violence at the municipal level (Nr=55). The variables taken into account are homicides, the number of guerrilla attacks, kidnapping, and displaced population. Miguel and Roland (Forthcoming) use district level army intelligence data (Nr=584) for assessing the long term socioeconomic impact of bombing in Vietnam. Li (2007) uses historic records on damage to 17 railroad lines across China to study the impact of China's warlord period on investment and economic growth.

Despite the increased attention for developing finer spatial measures of conflict intensity, there is still much room for improvement. In particular, there is a lack of uniformity in the sense that the fine micro-level conflict intensity measures used so far are based on rather particular country-specific data. A noteworthy effort made for developing a uniform micro-level measure of violent conflict intensity is the Armed Conflict Location and Event Data (ACLED, Raleigh, Linke, Hegre, and Karisen (2010)), which provides geo-referenced information on the location of battles and military activity. This information is derived by screening news articles with language recognition software. However, press accounts are often biased in the rapportation of events and computer news screening is sensitive to the language in which events are reported. Another drawback, highlighted by Restrepo, Spagat, and Vargas (2006), is that battle events (e.g. between a rebel group and government troops) may come short in reflecting the conflict's impact on civilians. These drawbacks can partly be overcome in time by using more sophisticated software programs and by coding different forms of violence.

A final set of type-II studies averages household level information at the community level to obtain a spatial conflict intensity measure (Bellows and Miguel (2009), Deininger (2003) and Shemyakina (Forthcoming)). This set of studies overlaps with the third and final type.

**Type III: Household conflict experience questions** Studies of type III rely on implicit or explicit household conflict experience questions for identifying the conflict's impact. Examples of implicit questions are those inquiring about refugee experience, damage to household dwellings and asset loss. A small number of such implicit household conflict experience questions are often part of commonly executed nationwide surveys (e.g. the Integrated Household Living Conditions Survey (IHLCS). Studies using such information include those that analyze socioeconomic outcomes across refugee and non refugee households (e.g. Kondylis (2008), Verwimp and Bavel (2004)), as well as one previously mentioned type-I study, Shemyakina (Forthcoming), which uses information on damage to household dwellings to study the impact of armed conflict in Tajikistan on schooling outcomes.

Explicit household conflict experience questions inquire about direct confrontations with violence, e.g. as a perpetrator or as a victim. Surveys with explicit conflict experience questions are often the result of the researchers' own effort and fieldwork, because such questions are rarely included in the usual nationwide surveys that have to pass through government institutions for their approval and implementation. This independence often pays off in detailed information from various conflict experience questions. The drawback of low profile - low budget surveys, is their small number of observations. Examples of type-III studies relying on explicit household conflict experience questions are Bellows and Miguel (2009), Deininger (2003), Verpoorten and Berlage (2007) and Verpoorten (2009). The latter two rely on a small scale survey and analyze the information at the household level while the former two use a nationally representative survey and aggregate the household answers at the community level in order to obtain a spatial measure of conflict intensity.

The concern for developing a more uniform measure of conflict intensity has triggered of a debate on the inclusion of a standard conflict module in household surveys (Brück, Justino, Verwimp, and Avdeenko (2010)). The introduction of such a module would most certainly be a way forward since it would provide detailed and comparable information on diverse conflict experiences at the level of households or individuals for a number of countries. However, in many post-conflict countries, it would take some time before a survey can be organized upon the restoration of peace, and the longer it takes, the more the data will be prone to recall bias as well as attrition bias which is especially large in post-conflict countries due to excess mortality and population displacement. While this drawback can be overcome in time with the help of funding, hand-on experience and innovative survey design, there are by now a large number of current post-conflict countries for which too much time will have past between the restoration of peace and the implementation of an ingenious post-conflict household survey. Besides, there is one drawback that cannot be overcome, i.e. the limited geographic coverage of a household survey. Even if the household survey is nationwide, it is often geographically stratified and many administrative units will not be covered. This limits the scope for matching the conflict intensity measure derived from the household data (e.g. by averaging household conflict experiences at the community level) with data from other sources, including other household surveys as well as statistics on local public and private investment.

## 3 A new measure of conflict intensity

This section develops a new spatial measure of conflict exposure that is not meant to replace existing measures, but is meant to be complementary and is particularly useful for measuring conflict cycles that are characterized by different forms of violence and a large death toll (direct or indirect). The basis for this measure is population census data, which is widely available. For example, among the 15 Sub-Saharan African countries experiencing civil war since the 1990s, 12 have had a post-conflict population census and two (Angola and DR Congo) are planning one in 2011 (UCDP/PRIO conflict data and UN population census statistics[4]). The exception is Somalia. Moreover, among the 12 countries with a post-conflict population census, we count 10 countries that have had a pre-conflict population census, which provides useful information for calculating a geographically disaggregated mortality baseline.

The measure is constructed in three steps. First, a number of mortality proxies are calculated separately for the pre- and post-war census at the level of the smallest administrative unit that they have in common[5]. These proxies may include widowhood (%), orphanhood

[4] http://unstats.un.org/unsd/demographic/sources/census/censusdates.htm

[5] Provided that this smallest common administrative unit is large enough in terms of population to reduce

(%) and mortality of sons and daughters (% of life births). For each of these proxies, age categories are defined. For example, the proportion of orphans may be calculated for individuals aged 30 or less. The choice of age limits is somewhat arbitrary and should be complemented with a sensitivity analysis.

We can write the pre- and post-war vectors of $p$ mortality proxies at the level of administrative unit $j$ as follows:

$$\begin{aligned} MP_j^{pre-war} &= [mp_{j1}, mp_{j2}, ..., mp_{jp}], \\ MP_j^{post-war} &= [mp'_{j1}, mp'_{j2}, ..., mp'_{jp}]. \end{aligned}$$

In a second step, the first difference is taken between the post-war and pre-war mortality proxies, yielding $j$ community level vectors of $p$ wartime excess mortality proxies ($wemp$):

$$MP_j^{post-war} - MP_j^{pre-war} = WEMP_j = [wemp_{j1}, wemp_{j2}, ..., wemp_{jp}].$$

Third, the set of wartime excess mortality proxies is summarized into an index by taking a weighted sum, with the weights determined by PCA[6]. The first PC will be an appropriate summary of excess mortality on a less to more scale if it captures a relatively high percentage of the total variance present in the excess mortality proxies set and the "loadings" of that PC have roughly equal values. The first PC, referred to as the Wartime Excess Mortality

---

the impact of outliers and erroneous data.

[6]PCA has the desirable property of reducing the dimensionality of a data set while retaining as much as possible of the variation present in the data set. From a set of variables, PCA extracts orthogonal linear combinations that capture the common information in the set most successfully. The first principal component (PC) identifies the linear combination of the variables with maximum variance, the second principal component yields a second linear combination of the variables, orthogonal to the first, with maximal remaining variance, and so on.

Formally, suppose that $x$ is a vector of $p$ random variables and $x^*$ is a vector of the standardized $p$ variables, having zero mean and unit variance, then the first principal component $PC1$ is the linear function $\alpha_1' x^*$ having maximum variance, where $\alpha_1$ is a vector of $p$ constants $\alpha_{11}, a_{12}, ..., \alpha_{1p}$ and $'$ denotes transpose.

$$PC1 = \alpha_1' x^* = \alpha_{11} x_1^* + \alpha_{12} x_2^* + ... + \alpha_{1p} x_p^*,$$

Mathematically, the vector $\alpha_1$ maximizes $var[\alpha_1' x^*] = \alpha_1' \Sigma \alpha_1$, with $\Sigma$ the covariance matrix of $x^*$,which corresponds to the correlation matrix of the vector $x$ of the original, unstandardized variables. For the purpose of finding a closed form solution for this maximization problem, a normalization constraint, $\alpha_1' \alpha_1 = 1$, is imposed. To maximize $\alpha_1' \Sigma \alpha_1$ subject to $\alpha_1' \alpha_1 = 1$, the standard approach is to use the technique of Lagrange multipliers. It can be shown that this maximization problem leads to choosing $\alpha_1$ as the eigenvector of $\Sigma$ corresponding to the largest eigenvalue of $\Sigma$, $\lambda_1$ and $var[\alpha_1' x^*] = \alpha_1' \Sigma \alpha_1 = \lambda_1$. To interpret the PC in terms of the original variables, each coefficient $\alpha_{1l}$ must be divided by the standard deviation, $s_l$, of the corresponding variable $x_l$. For example, a one unit increase in $x_l$, leads to a change in the $1st$ PC equal to $\alpha_{1l}/s_l$.For a detailed exposition of principal component analysis we refer to Jolliffe (2002) and Dunteman (2001).

Index, can be written as:

$$WEMI_j = l \times WEMP_j = l \times [wemp_{j1}, wemp_{j2}, ..., wemp_{jp}]$$

with $l$ the vector of loadings obtained through PCA.

A number of studies have used PCA for the purpose of summarizing conflict indicators by a conflict index. Pioneering work by Hibbs (1973) derives indices of "collective protest" and "internal war" from a 108-nation cross-sectional analysis of six event variables on mass political violence. Following Hibbs (1973) a large number of cross-country studies have used an index of sociopolitical instability as an explanatory variable in regressions in which the dependent variable is growth, savings or investment (e.g. Venieris and Gupta (1986), Barro (1991), Alesina and Perotti (1996)). To the best of our knowledge, only one micro-economic study, González and Lopez (2007), uses PCA to summarize variables into a micro level index of violent conflict. This study, looking at the effect of political violence in Columbia on farm household efficiency, defines five indicators of violence: homicides, the number of attacks by FARC guerrillas, the number of attacks by ELN guerrillas, kidnappings, and displaced population. The first PC accounts for 43% of the joint variance of the five indicators, and is retained as an index of political violence. The main difference between these previous studies and the measure proposed in this study is that $WEMI$ relies on population census data instead of event data or data from transitional justice records.

In the demographic literature, there is a long tradition of studies that infer excess mortality from characteristics of the surviving population, a method referred to as indirect mortality estimation (e.g. Hill and Trussell (1977), Timaeus (1986)). The main difference with the current approach is that, instead of trying to come up with an absolute number for excess mortality, which is far more demanding in terms of data requirements and assumptions, the $WEMI$ aims at capturing relative excess mortality, i.e. its spatial distribution on a less to more scale.

# 4 Illustration: measuring the Rwandan conflict cycle

## 4.1 The Rwandan conflict cycle

The Rwandan conflict cycle of the nineties included civil war, genocide, reprisal killings, (counter)insurgency and a major refugee crisis. Civil war broke out in Rwanda at the end of 1990, when the RPF (Rwandan Patriotic Front), a rebel army consisting mostly of Tutsi exiles, started launching attacks from Uganda[7]. Intermittent hostilities and negotiations between the FAR (Armed Forces of Rwanda) and the RPF resulted in a power sharing agreement, signed in Arusha (Tanzania) on August 4, 1993. But on April 6, 1994 the plane carrying President Habyarimana of Rwanda on his way back from negotiations in Dar-es-Salam was shot down while landing. Thereafter, Rwanda descended into chaos. Within hours, the FAR, the *Interahamwe*[8], administrators and ordinary people started to kill Tutsi, moderate Hutu and Hutu leaders from political parties rival to the president's party, the MRND (National Revolutionary Movement for Development). Simultaneously with the start of genocide (and politicide), the civil war between the FAR and the RPF resumed.

Late in June 1994, the RPF took power and the massive killings came to an end. The balance of the events was shocking: an estimated 800,000 Tutsi and moderate Hutu killed in a time span of barely 100 days ( Des Forges (1999), Prunier (1998), Verpoorten (2005)). In addition, tens of thousands of people died from deprivation in refugee camps, became direct or indirect victims of the civil war between the RPF and the FAR, or fell victim to reprisal killings by the RPF both inside Rwanda and across the border with Congo (Davenport and Stam (2009), Reyntjens (2009), the "Gersony report"[9], Verpoorten (2010a)). Regarding the two latter forms of violence, Davenport and Stam (2009) estimate that, during April-June 1994, the sum of victims in zones under RPF control and the zones contested by the RPF and FAR amounted to respectively 80,000 and 90,000.

Even when relative peace was established in Rwanda by the end of 1994, until the late nineties, violence continued in the Northwest which was the locus of insurgency and counterinsurgency operations on the part of respectively the FAR and Interhamwe - who used Congo as a basis for sporadic attacks, and the RPF - who remained military active close to and across the border with Congo in order to fight the insurgents.

While these events all occurred in the nineties, their geographic location within Rwanda

[7] The population of Rwanda and of the neighboring Burundi consists of three groups: Hutu, Tutsi and a small minority of Twa. Before the genocide the Tutsi made up approximately 10% of the population and the Twa 1%. The remaining population were Hutu. Before the colonization the borderline between Hutu and Tutsi was quite flexible, because at times families could change status. Moreover, a substantial number of chiefs were Hutu, especially those supervising access to land. Under Belgian colonial rule the distinction between the two groups became more rigid. Almost all chiefs were Tutsi (See Prunier, 1995, chapter 1). In 1959 while the country was still under Belgian rule, the Tutsi king was removed and Hutu took over the government.

[8] *Interahamwe literally means "those who stand together" or "those who attack together". This militia was formed by Habyarimana's political party in 1992, when the party started giving military training to its youth.*

[9] The "Gersony Report" is the name given to an unpublished report that identified a pattern of massacres by the RPF. The findings in the report were made by a team under Robert Gersony under contract to the United Nations High Commissioner for Refugees. Gersony's personal conclusion was that between April and August 1994, the RPF had killed "between 25,000 and 45,000 persons, between 5,000 and 10,000 persons each month from April through July and 5,000 for the month of August" (Des Forges (1999)).

differed, with the 1991-1993 civil war confined to the northern provinces, the 1994 genocide especially severe in the South of the country, the 1994 civil war most intense in and around the capital city and the 1994-1998 (counter)insurgency concentrated in the Northwest (see Figure 1). Studies discussing the spatial pattern of the conflict cycle include Des Forges (1999), Davenport and Stam (2009) and Verpoorten (2010a).

---

Insert Fig. 1 about here

---

### 4.2 The Wartime Excess Mortality Index (WEMI)

For the construction of $WEMI$, I rely on a 10% random draw of the 1991 population census (N=742,918 individuals; accessed online from IPUMS International) and the entire 2002 population census (N=8.1 million; received from the Rwandan Government). Verpoorten (2005) demonstrates that the Rwandan census data are very reliable except for the recording of ethnicity[10]. The 2002 census includes information at the level of the administrative sectors (N=1540). However, the smallest available administrative unit in the 1991 population census is the commune (N=145), which is one administrative unit above the sector. Hence, I calculate the excess mortality proxies at the commune level, which still is a fairly small administrative unit with an average size of 174 squared km.

An important note to make is that the aftermath of the Rwandan conflict was characterized by a huge refugee crisis and considerable internal and external migration. Therefore, I calculate the commune level 2002 mortality proxies ($mp$) both including and excluding individuals who changed residence over the period 1991-2002. The results excluding migrants are presented in this article. The results including migrants are qualitatively the same and can be obtained on request.

Taking the difference between 2002 and 1991 $mp$, I derive the following five wartime excess mortality proxies ($wemp$) for 145 communes ($j = 1...145$) in Rwanda:

- ($wemp_{j1}$) ΔMortality of sons : 2002-1991 difference in total number of boys died/number of boys born (for all women who ever gave birth);
- ($wemp_{j2}$) ΔMortality of daughters : 2002-1991 difference in total number of girls died/girls born (for all women who ever gave birth);

[10] Tutsi were underreported in the census, either by the Habyarimana regime to keep their school enrolment and public employment quotas low, either by Tutsi themselves to avoid discimination.

- ($wemp_{j3}$) $\Delta$Widowhood : 2002-1991 difference in the proportion of widows among women (>=30 years);

- ($wemp_{j4}$) $\Delta$Double orphanhood : 2002-1991 difference in the proportion of double orphans among children and youngsters (<30 years);

- ($wemp_{j5}$) Disability : the proportion of the 2002 population reporting a handicap due to war or genocide.

Table 2 provides summary statistics for the pre-war and post-war *mp* and their first differences (*wemp*). For all *mp*, we find higher values in the 2002 population census than in the 1991 population census. The mortality of sons and daughters increased with respectively 8 (from 20% to 28%) and 7 (from 18% to 24%) percentage points. Widowhood rose from 18% to 31% and orphanhood from 2% to 5%. Disability due to genocide or war was not present in the 1991 census and amounted to 0.3% in 2002.

Subjecting the set ($wemp_{j1} - wemp_{j5}$) to PCA results in the following first PC:

$$WEMI_j = 0.48 \times wemp_{j1} + 0.43 \times wemp_{j2} + 0.46 \times wemp_{j3}$$

$$+0.48 \times wemp_{j4} + 0.38 \times wemp_{j5},$$

which explains up to 69% of the total variation of the variable set and has significant positive loadings on all Wartime Excess Mortality Proxies.

Figure 2a plots the quintiles of $WEMI_j$ on an administrative map of Rwanda. We observe a large number of top quintile communes in Butare, in and around Kigali City, in Gisenyi, in the northwestern corner of Kibungo and in the southwestern corner of Umutara. In addition, we find smaller local clusters in the West of Ruhengeri and Kibuye and in the Southeast of Gitarama and Gikongoro. Taken together, this map reflects well the different forms of violence in Rwanda based on event data. For a detailed analysis of spatial determinants of $WEMI$, I refer to Verpoorten (2010a) who demonstrates that the pattern of $WEMI$ is strongly related to spatial determinants of the 1994 genocide, the 1994 open warfare between the RPF and the FAR, the 1994-1998 refugee crisis and the 1995-1998 (counter)insurgency.

---

Insert Table 2 about here

---

---

Insert Figure 2 about here

---

### 4.3 Genocide Excess Mortality and other Excess Mortality (GEMI, CEMI)

I develop a two-dimensional conflict intensity measure to distinguish between genocide and other forms of violence. In order to do so, I expand the set of variables subjected to PCA with rich information on genocide intensity. This information is taken from the transitional Rwandan justice system in charge of judging 1994 genocide suspects (gacaca), which collected information on genocide suspects and survivors in the course of 2005-2007 (Government of Rwanda (2005)). The numbers of genocide suspects and survivors were made public in the course of 2007. Verpoorten (2010b) discusses the reliability of the data and explains in detail how the data can be transformed to obtain genocide intensity proxies.

The six Genocide Proxies that can be derived from the gacaca data include the number of genocide suspects classified in three categories ($gp_1 - gp_3$) as well as three categories of surviving genocide victims ($gp_4 - gp_6$), i.e. close relatives of persons who were killed during the genocide. Defined at the commune level $j$ and taken proportional to the 1994 population, this results in the following six variables:

- ($gp_{j1}$) Genocide suspects accused of planning, organizing or supervising the genocide, and committing sexual torture (% 1994 population)
- ($gp_{j2}$) Genocide suspects accused of killings or other serious physical assaults (% 1994 population)
- ($gp_{j3}$) Genocide suspects accused of looting or other offences against property (% 1994 population)
- ($gp_{j4}$) Widowed genocide survivors (% 1994 population)
- ($gp_{j5}$) Orphaned genocide survivors (% 1994 population)
- ($gp_{j6}$) Disabled genocide survivors (% 1994 population)

Table 3 provides summary statistics. The genocide suspects of category 1, 2 and 3 account on average for 1.2%, 6.8% and 5.0% of the 1994 population. The categories of genocide survivors - widows, orphans and disabled - make up a smaller proportion of the population, respectively 0.5%, 1.2% and 0.2%.

The six genocide proxies ($gp_{j1} - gp_{j6}$) and the five wartime excess mortality proxies ($wemp_{j1} - wemp_{j5}$) are subjected to PCA to obtain up to 11 PCs. Denoting the latter by the vector $e_j$ and the former by the vector $g_j$, the first four PCs are given by the following linear combinations:

$$
\begin{aligned}
PC1_j &= [0.23, 0.19, 0.12, 0.16, 0.23]' \times e_j + [0.39, 0.39, 0.31, 0.39, 0.39, 0.33]' \times g_j \\
PC2_j &= [0.40, 0.38, 0.46, 0.46, 0.31]' \times e_j + [-0.15, -0.19, -0.25, -0.18, -0.15, -0.08]' \times g_j \\
PC3_j &= [0.20, 0.33, 0.04, 0.02, -0.49]' \times e_j + [0.19, 0.27, 0.45, -0.24, -0.20, -0.45]' \times g_j \\
PC4_j &= [-0.33, -0.57, 0.39, 0.40, 0.15]' \times e_j + [0.25, 0.02, 0.32, -0.19, -0.19, -0.06]' \times g_j
\end{aligned}
$$

Combined, these four PCs explain 86% of the total variance in the underlying set of 11 variables. However, PC3 and PC4 add relatively little to the explained variance, respectively 9% and 5%, and they do not suggest a clear interpretation. In contrast, the first two principal components each explain a sizeable proportion of the variance (43% and 29% respectively) and combined account for 72% of the joint variance of the 11 variables. Moreover, they both have a rather straightforward interpretation.

The loadings for the first PC are highest for the genocide proxies, $g_j$, while the excess mortality proxies, $e_j$, are the dominant group in the second PC. Given that the first PC has dominant loadings on the genocide proxies, it is interpreted as a Genocide Excess Mortality Index ($GEMI$). The loadings on the second PC imply that, after genocide related violence has been accounted for, the main source of variation is between sectors with large excess mortality proxies relative to the genocide proxies. Therefore, the second principal component is interpreted as excess mortality stemming from other sources, including Civil war, reprisal killings, the refugee Crisis and (Counter)insurgency ($CEMI$)[11].

Figures 2b and 2c plot quintiles of respectively $GEMI$ and $CEMI$. We find the highest values for $GEMI$ in the South of Rwanda, in the center close to Kigali City and in the eastern province Kibungo. This pattern corresponds with the event data underlying Figure 1 as well as with the intensity of genocide described in other studies (Davenport and Stam (2009), Justino and Verwimp (2006), Verpoorten (2010b)). $CEMI$ yields the highest values in the northern provinces of Gisenyi and Ruhengeri, around Kigali City as well as in the eastern provinces Byumba and Kibungo. This spatial pattern coincides with the location of RPF battle fronts and zones that were situated in the eastern and central part of Rwanda

[11] See Jolliffe (2002) for a comprehensive exposition on the interpretation of PCs.

in the course of the months April-July 1994, and finally, in the period 1995-1998, in the Northwest at the border with Congo (Davenport and Stam (2009), Verpoorten (2010a)).

---

Insert Table 3 about here

---

### 4.4 Accounting for possible survival bias

The conflict intensity indices, $WEMI$, $GEMI$ and $CEMI$ may be biased downward in communes where many families were entirely exterminated. In order to attenuate the effect of survival bias, I increase the weight of communes that are close to sites of large-scale massacres. The proximity to a large-scale massacre is taken into account by adding the natural logarithm of the commune level distance to the nearest mass grave to the set of variables subjected to PCA. This distance is calculated in km by overlaying a geo-referenced administrative map with the location of 71 mass graves in Rwanda taken from the Yale Genocide Studies website.

The resulting conflict indices are given by the following linear combinations:

$$\begin{aligned} WEMI\_s_j &= [0.47, 0.42, 0.45, 0.48, 0.38]' \times e_j + [-0.15] \times s_j \\ GEMI\_s_j &= [0.21, 0.16, 0.11, 0.15, 0.21]' \times e_j + \\ &\quad [0.38, 0.38, 0.30, 0.38, 0.38, 0.31]' \times g_j + [-0.30] \times s_j \\ CEMI\_s_j &= [0.41, 0.39, 0.46, 0.46, 0.31]' \times e_j \\ &\quad +[-0.13, -0.17, -0.24, -0.16, -0.13, -0.07]' \times g_i + [0.04] \times s_j \end{aligned}$$

with $s_j$ "log(distance to mass grave)" and $g_j$ and $e_j$ as defined above.

This adjustment for survival bias does not yield qualitatively different results with respect to the spatial pattern of the indices (Figure 2) or in the application of armed conflict and schooling (Section 5). A sensitivity analysis with respect to the age limits for the excess mortality proxies $wemp_{j3}$ and $wemp_{j4}$ (setting the age limits 5 years lower (at 25) or 5 years higher (at 35)) neither alters the qualitative results. The results of these robustness checks are not reported in this paper, but can be obtained on request.

# 5 Application: armed conflict and schooling in Rwanda

The aim of the application is to assess how the newly proposed one- and two-dimensional measures perform in an empirical application. Evaluating the usefulness of a measure by means of an empirical application is problematic when the expected outcome is contested by theory or by previous empirical studies. Hence, the application focuses on a well-established impact of armed conflict, i.e. its negative impact on schooling (e.g. Lai and Thyne (2007)). The section proceeds in three steps. First, I describe the data on schooling. Second, I use conflict exposure in time to identify the impact of armed conflict on schooling with a difference-in-difference (DD) estimate, and I do so for all 145 communes separately, obtaining a spatial pattern of the 1991-2002 schooling deficit, which is compared with the spatial pattern of $WEMI$, $GEMI$ and $CEMI$. Third, I combine information on conflict exposure in time with conflict exposure in space to calculate (both OLS and IV) difference-in-difference-in-difference (DDD) estimates that relate the 1991-2002 schooling deficit to $WEMI$, $GEMI$ and $CEMI$.

## 5.1 Schooling data

I use information on the number of years of schooling for individuals aged 6 to 50 in a 10% random draw from the 1991 and 2002 population census[12]. The individuals are divided across a young and an old cohort. The young age cohort in 2002 represents the group of individuals exposed to the armed conflict at primary schooling age (6-12). The age limits of this age cohort are set at 6 and 22 since 6 is the age at which children start primary school and those aged 22 in 2002 were aged 12 at the start of the conflict in 1990 (alternative age categories give similar results - see subsection 5.4 for a discussion).

In the 10% random draw of the 1991 and 2002 census, the young age cohort (6-22) counts respectively 305,881 and 347,540 individuals, while the old age cohort (23-50) counts respectively 211,007 and 221,025 individuals. Table 4 provides summary statistics on schooling attainment across 1991 and 2002 for both age groups combined. The figures show an increase over time in the proportion of individuals with some primary schooling (from 55.5% to 61.1%) and with some secondary schooling (from 5.1% to 8.2%).

However, when distinguishing between the age groups, we find that this progress in schooling is entirely driven by the old age cohort. Table 5 shows a progress of 0.8 years of

[12] I use only 10% of the 2002 census, because STATA cannot perform the regression analysis for the entire census (8.1 million observations).

schooling for the old age cohort and a drop of 0.2 years for the young age cohort, yielding a DD estimate of almost one year of schooling, or, when taking log(years of schooling) as a dependent variable, a 20% decrease in the number of years of schooling.

This result is in line with the study of Akresh and de Walque (2008) who study years of education of a young (6-15) and an older (16-35) age cohort in the 1992 and 2000 DHS survey and conclude that children exposed to armed conflict completed close to one-half year less education which corresponds to a 18.3% drop relative to the average educational achievement.

---

Insert Table 4 about here

---

Insert Table 5 about here

---

### 5.2 Difference-in-Difference estimates: the spatial pattern of the 1991-2002 schooling deficit

Given that the data used for this analysis include 1,074,561 individuals divided across 145 communes, we have sufficient observations to calculate commune level DD estimates to identify the spatial pattern of the schooling deficit. The commune level DD are obtained by estimating the following equation for each commune $j$ separately:

$$Y_{ijt} = \alpha_{j0} + \alpha_{j1}(T_t \times young_i) + \alpha_{j2}T_t + \alpha_{j3}young_i + \varepsilon^1_{ijt}$$

WITH

$Y_{ijt}$ : average years of education of individual aged $i$ in commune $j$ at time $t$

$T_t$ : indicator variable for being in the 2002 census

$young_i$ : indicator variable for being in the young age cohort

$\varepsilon_{ijt}$ : idiosyncratic error

The coefficients $\alpha_{j1}$ give estimates of the 1991-2002 schooling deficit for 145 different communes. It are DD estimates identified from discrete treatment $T_t$, with the young age cohort the treated and the old age cohort the non-treated group. Figure 3 plots the quintiles of $\hat{\alpha}_{j1}$ on a map. The spatial pattern shows clusters of large drops in schooling in the Northwest, scattered throughout the centre, the South and East. While this pattern is very different from $GEMI$ (Fig. 2b), it is similar to $WEMI$ (Fig. 2a), which suggests that the 1991-2002 schooling deficit in Rwanda is not only due to genocide, but also to other events of the conflict cycle.

---

Insert Fig. 3 about here

---

### 5.3 Difference-in-Difference-in-Difference estimates: explaining the 1991-2002 schooling deficit

Finding a close match between two spatial patterns is not enough to establish causality between armed conflict and the schooling deficit. To determine causality, we take the analysis two steps further. First, we estimate a DDD effect in which treatment is defined as being in the young age cohort in the 2002 sample and residing in a geographic unit with high violent conflict intensity. Second, we instrument for conflict intensity.

The DDD estimate corresponds to the coefficient $\beta_1$ in the following equation:

$$\begin{aligned} Y_{ijt} &= \beta_0 + \beta_1(T_t \times C_j \times young_i) + \beta_2(T_t \times C_j) + \beta_3(T_t \times young_i) \\ &\quad + \beta_4(C_j \times young_i) + \beta_5 T_t + \beta_6 C_j + \beta_7 X_{ijt} + \zeta_i^2 + \eta_j^2 + \varepsilon_{ijt}^2 \end{aligned}$$

WITH

$Y_{ijt}$ : average years of education of individual $i$ in commune $j$ at time $t$

$T_t$ : indicator variable for being in the 2002 census

$C_j$ : commune level measure of conflict intensity, rescaled to fit the interval $[0, 1]$

$young_i$ : indicator variable for being in the young age cohort

$X_{ijt}$ : household and individual level controls[13]

$\zeta_i$ : age fixed effects

$\eta_j$ : province fixed effects

$\varepsilon_{ijt}$ : idiosyncratic error

---

Insert Table 6 about here

---

Table 6 gives the OLS results with $C_j$ defined as the previously derived conflict indices $WEMI_j$, $GEMI_j$ and $CEMI_j$, respectively in columns (1), (2) and (3). The estimated coefficient $\hat{\beta}_1$is negative and highly significant, with respective values of $-0.59, -0.13$ and $-0.63$. Since the conflict indices are rescaled to the interval $[0, 1]$, these values can be interpreted as the change in the number of years of education when moving from zero to maximum conflict intensity. These results suggest that, compared to $GEMI_j$, the impact of $CEMI_j$ on schooling is more severe, which stands in contrast with the study of Akresh and de Walque (2008) who attribute the entire estimated schooling deficit to genocide. In column (4), $GEMI_j$ and $CEMI_j$ are jointly included, which yields very similar point estimates to those of columns (2) and (3) because $GEMI_j$ and $CEMI_j$ are orthogonal vectors (which is a feature of PCA).

The estimates for $\beta_1$ may be biased due to reversed causality, i.e. the incidence or intensity of conflict may have been higher in those areas where there was a downward trend in education. Alternatively, there may be omitted variable bias, i.e. excess mortality may have increased for another reason besides conflict and this increase may have gone hand in hand with a decrease in education. Both of these sources of endogeneity bias can be remedied for with an instrumental variable approach.

We use the commune level distance to Uganda as a first identifying instrument. At the peak of the civil war and genocide in 1994 the RPF infiltrated from Uganda and gradually moved towards Kigali City engaging in heavy battles with the Rwandan army before eventually taking over the capital. The battle front then moved to other areas of the country, safeguarding the remaining Tutsi from being killed and engaging in reprisal killings on Hutu who allegedly participated in the genocide (Davenport and Stam (2009))[14]. The exogenous character of the border stems from colonial history, as it was fixed following a compromise agreement by European nations[15].

A second identifying instrument for conflict intensity is the commune level distance to Nyanza, a sector located in the northwestern corner of Butare, close to the border with

[14] Previous work on Rwanda and elsewhere has used similar instruments. Miguel and Roland (Forthcoming) use the distance to the meridian that distinguishes North from South Vietnam as an instrument for bombing intensity in Vietnam, while Akresh and de Walque (2008) also use distance to Uganda, measured at the level of 11 provinces, as identifying instrument for armed conflict in Rwanda. In this study I use more detailed spatial data, and use the distance of each of the 145 communes to Uganda. The use of this finer instrument should considerably add to its strength.

[15] The border was fixed in 1910 at an Anglo - German - Belgian Convention in Brussels, following a 1890 compromise agreement between the United Kingdom and Germany that provided that the boundary between British and German territories should follow the parallel of one degree south latitude across Lake Victoria and should continue westward to its intersection with the 30th meridian of longitude east of Greenwich (Department of State (1965)).

Gikongoro and Gitarama province. Nyanza was the capital of the Tutsi monarchy, which controlled most of the present-day Rwandan territory from as early as the 14th century. Its economic and political importance faded during colonization and abruptly ended with the 1959 revolution and subsequent 1962 independence. The government reorganized Rwanda's administrative division shortly after independence. The southern and western outskirts of the Nyanza region were attached to what is now the eastern part of Gikongoro, a highland area inhabited largely by Hutu. The aim was to weaken Tutsi influence around the former royal capital Nyanza (Des Forges (1999)). Today nothing is left of Nyanza's former glory, but the proportion of Tutsi in the communities close to Nyanza was still higher prior to the genocide (Verpoorten (2005)), which makes it a relevant instrument for the intensity of ethnic violence.

The first stage IV results are reported in Table 8 and demonstrate the relevance of both instruments[16]. Table 7 reports the second stage IV results[17]. Compared to the OLS results, the IV results for $\beta_1$ in columns (1), (3) and (4) are very similar. However the IV estimate in column (2), which measures the impact of $\widehat{GEMI}_j$, is not significantly different from zero. This is likely due to the failure to control for other forms of violence, which now affect the control group of the DD. Note that this is not an issue in the OLS results in Table 6 because $GEMI_j$ and $CEMI_j$ are orthogonal vectors. However, upon instrumenting, the orthogonality between $GEMI_j$ and $CEMI_j$ disappears. When adding $\widehat{CEMI}_j$ as a control variable in column (4), the coefficient on $\widehat{GEMI}_j$ turns again negative and significant. These results suggest that failing to take into account all forms of violence may result in severely biased estimates.

—————————————-

Insert Table 7&8 about here

—————————————-

## 5.4 Robustness checks

Both the OLS and IV estimate for $\beta_1$ remain qualitatively the same after correcting for survival bias, i.e. when replacing the conflict indices by $WEMI\_s_j$, $GEMI\_s_j$ and $CEMI\_s_j$

[16] Distance to Nyanza has the expected negative sign, reflecting high genocide intensity near Nyanza. Distance to Uganda has a positive sign, indicating high conflict intensity further away from the border with Uganda, reflecting the 1994 civil war in the centre and the east, with the battle intensifying as the RPF proceeded.

[17] Because we estimate the IV models with 2SLS, which yields consistent parameter estimates but biased standard errors, we correct the standard errors using the procedure proposed by Baltagi (2002).

(results reported in Table 9). Looking at the quantitative effect, we find that the estimated OLS (IV) coefficient on $GEMI\_s_j$ is 13.8% (35.4%) larger compared to the one on $GEMI_j$, whereas the difference is negligibly small for $CEMI_j$ and $CEMI\_s_j$. This suggests that survival bias is especially an issue for the measurement of genocide related excess mortality.

I run three other robustness checks. First, I use $WEMI\_m_j$, $GEMI\_m_j$ and $CEMI\_m_j$ which are the conflict indices calculated for the entire 2002 population, i.e. including migrants. Second, I use different age categories for widowhood and orphanhood (-5 years and +5 years) to calculate $wemp_{j3}$ and $wemp_{j4}$. Finally, I replicate the results using different definitions for the young and old age cohort, including those used by Akresh and de Walque (2008), i.e. 6-15 and 16-35, as well as 8-17 and 17-37. All robustness checks yield results that are qualitatively the same as those presented in Tables 4 to 8, and they can be obtained on request.

---

Insert Table 9 about here

---

## 6 Comparison with previous measures

Table 10 lists the results of the OLS and IV estimates of the coefficient $\beta_1$ for cases where $C_j$ equals conflict intensity measures that were described and/or used in previous studies. A detailed description of all 15 alternative measures can be consulted in the Appendix.

The 15 measures include (i) *five dummies of conflict intensity at the province level* previously used in studies by Justino and Verwimp (2006), Akresh, Verwimp, and Bundervoet (forthcoming) and Akresh and de Walque (2008); (ii) *six other province level measures* derived from diverse sources including the Genodynamics Project (Davenport and Stam (2009)), Yale Genocide Studies, the 1991 population census, ACLED (Raleigh, Linke, Hegre, and Karisen (2010)) and the records of the transitional justice system (Government of Rwanda (2005)); (iii) *four commune level measures* based on data from Yale Genocide Studies, the 1991 population census and the gacaca records (previously used by Yanagizawa (2010)).

The OLS results for these alternative measures are mixed: (i) three of the five province level dummies yield a significantly negative estimate for $\beta_1$, while two result in a significantly positive estimate; (ii) among the six continuous province level measures, two yield a significantly positive estimate for $\beta_1$, and four give an insignificant result; (iii) three out of the four commune level measures result in the expected effect on schooling, and the remaining one gives an insignificant result. The IV results are also very different across measures and can be consulted in Table 10.

Hence, even in a seemingly straightforward application of the impact of armed conflict on schooling, different micro-level conflict measures yield strinkingly different results.

Insert Table 10 about here

## 7 Conclusion

How should we measure micro-level conflict intensity in order to take a step forward in the study of the legacy of armed conflict? This paper argues that there is a need for finer measures that can capture different forms of violence and that can be used to compare results across countries.

It is demonstrated how widely available data, i.e. population census data, and a commonly used method, Principal Component Analysis, can result in a fine index of Wartime Excess Mortality, referred to as WEMI. This index is neutral to the cause of excess mortality and is therefore well suited in a context in which different forms of violence took place.

Its neutrality may be a drawback in the sense that $WEMI$ may capture other causes of mortality, besides the direct and indirect impact of violence. However, this drawback can be taken care of with an instrumental variable approach, which has become standard practice by now in empirical studies on the impact of armed conflict to correct for the possible endogeneity of conflict intensity.

The $WEMI$ is calculated for the case of Rwanda and it is shown that its spatial pattern corresponds well with three main forms of violence taking place in Rwanda in the nineties: civil war, genocide and (counter-)insurgency. Moreover, adding detailed data on the genocide to the set of "neutral" excess mortality proxies, it is demonstrated how one can distinguish between genocide excess mortality and other forms of excess mortality by means of a two-dimensional measure ($GEMI$, $CEMI$).

Both the one- and two-dimensional index indicate a significantly negative impact of armed conflict on schooling in Rwanda. In contrast to Akresh and de Walque (2008) who attribute the entire schooling deficit in the nineties to genocide, the findings in this paper show that other events in the Rwandan conflict cycle explain at least half of the schooling deficit.

The proposed one- and two-dimensional measure can contribute to the micro-economic study of the legacy of different forms of violence in Rwanda. On the one hand, previous studies on health and educational outcomes can be replicated with a richer measure of violent conflict intensity, allowing deeper understanding of e.g. the impact of different forms of violence. On the other hand, new issues in the area of technology, institutions and social norms that require fine spatial information on conflict intensity can be addressed.

The $WEMI$ can be calculated for a set of other post-conflict countries, which may open the perspective for useful comparisons of country case studies. Admittedly, since WEMI is a relative rather than an absolute measure (i.e. excess mortality on a less to more scale across administrative units of a single country), the results of different country case studies need to be interpreted accordingly. In addition, the set of countries for which WEMI can be calculated will be a nonrandom subset of post-conflict countries, i.e. countries where violence resulted in high excess mortality and where post-conflict institutions are sufficiently (co)operative to collect and release population census data.

In countries where different forms of violence took place and rich data is available on at least one form of violence, the WEMI can be a basis for developing a multi-dimensional index. If we succeed in identifying how different forms of violence within the borders of the same country have different effects on the socioeconomic variables studied, we may come a long way in understanding the impact of violent conflict on poverty.

# 8 Acknowledgments

I'm grateful to Tilman Brück, Geert Dhaene, Koen Decancq, Scott Gates, Nils Petter Gleditsch, Romain Houssa, Pablo Rovira Kaltwasser, Håvard Mokleiv Nygård, Pieter Serneels, Håvard Strand, Henrik Urdal, Bruno Versailles and participants at seminars and conferences at universities in Leuven (LICOS seminar), Amsterdam (Jan Tinbergen Conference), Namur (LICOS/CRED seminar), Oxford (CSAE conference), Oslo (PRIO) and Berlin (HiCN & DIW Berlin Conference) for very helpful comments. I owe thanks to the Rwandan National Service of Gacaca Jurisdiction, the Rwandan Geographic Information System Centre, the Rwandan National Census Service and IPUMS International for making available the data used in this study. All errors and opinions expressed remain my own.

## References


Akresh, R., and D. de Walque (2008): "Armed Conflict and Schooling: Evidence from the 1994 Rwandan Genocide," HiCN Working Papers 47, Households in Conflict Network.

Akresh, R., P. Verwimp, and T. Bundervoet (forthcoming): "Civil war, crop failure, and child stunting in Rwanda," Economic Development and Cultural Change, (4208).

Alderman, H., J. Hoddinott, and B. Kinsey (2006): "Long term consequences of early childhood malnutrition," Oxford Economic Papers, 58(3), 450–474.

Alesina, A., and R. Perotti (1996): "Income distribution, political instability, and investment," European Economic Review, 40(6), 1203–1228.

Barro, R. J. (1991): "Economic growth in a cross-section of countries," Quarterly Journal of Economics, 106, 407–444.

Bellows, J., and E. Miguel (2009): "War and Institutions: New Evidence from Sierra Leone," American Economic Review, 96(2), 394–399.

Blattman, C., and E. Miguel (2010): "Civil War," Journal of Economic Literature, 48(1), 3–57.

Brück, T., P. Justino, P. Verwimp, and A. Avdeenko (2010): "Identifying Conflict and Violence in Micro-Level Surveys," Discussion paper, IZA DP No. 5067.

Bundervoet, T. (2007): "Livestock, Activity Choices and Conflict: Evidence from Burundi," HiCN Working Paper 24, Households in Conflict Network.

Bundervoet, T., P. Verwimp, and R. Akresh (2009): "Health and Civil War in Rural Burundi," Journal of Human Resources, 44(2), 536–563.

Chamarbagwala, R., and H. E. Morán (2008): "The Human Capital Consequences of Civil War: Evidence from Guatemala," HiCN Working Paper 59, Households in Conflict Network.

Davenport, C., and A. Stam (2009): "Rwandan Political Violence in Space and Time," Unpublished manuscript.

de Walque, D. (2004): "The long-term legacy of the Khmer Rouge period in Cambodia," Policy Research Working Paper Series 3446, The World Bank.

Deininger, K. (2003): "Causes and consequences of civil strife: micro-level evidence from Uganda," Oxford Economic Papers, 55(4), 579–606.

Department of State, . (1965): "International Boundary Study," Discussion paper, Washington, D.C., US.

Des Forges, A. (1999): Leave None to Tell the Story: Genocide in Rwanda. New York: Human Rights Watch.

Dunteman, G. H. (2001): Principal Components Analysis, Sage university papers, Quantitative applictions in the social sciences. Sage.

Gleditsch, N. P., P. Wallensteen, M. Eriksson, M. Sollenberg, and H. Strand (2002): "Armed Conflict 1946–2001: A New Dataset," Journal of Peace Research, 39(5), 651–637.

González, M. A., and R. A. Lopez (2007): "Political Violence and Farm Household Efficiency in Colombia.," Economic Development and Cultural Change, 55(2), 367 – 392.

Government of Rwanda, . (2005): "Report on Data Collection: Annexes," Discussion paper, National service of Gacaca Jurisdiction.

Hibbs, D. (1973): Mass Political Violence: A Cross-national Causal Analysis. John Wiley & Sons.

Hill, K., and T. Trussell (1977): "Further Developments in Indirect Mortality Estimation," Population Studies, 31, 75–84.

Human Security Report Project, . (2010): "Human Security Report 2009/2010: The Causes of Peace and the Shrinking Costs of War. Pre-publication.," Discussion paper, Forthcoming in print from Oxford University Press.

Jolliffe, I. (2002): Principal Component Analysis, no. XXIX in Springer Series in Statistics. Springer, NY, 2nd edn.

Justino, P., and P. Verwimp (2006): "Poverty Dynamics, Violent Conflict and Convergence in Rwanda," HiCN Working Papers 16, Households in Conflict Network.

Kondylis, F. (2008): "Agricultural Outputs and Conflict Displacement: Evidence from a Policy Intervention in Rwanda," Economic Development and Cultural Change, 57(1), 31–66.

Lai, B., and C. Thyne (2007): "The effect of civil war on education, 1980-97," Journal of Peace Research, 44(3), 277–292.

Leon, G. (2010): "Civil Conflict and Human Capital Accumulation: The Long Term Effects of Political Violence in Peru," Working Papers id:2505, esocialsciences.com.

Li, D. (2007): "The Economic effects of Civil War: Evidence from Chinese National Railroads, 1906-1923," Unpublished manuscript.

Miguel, E., and G. Roland (Forthcoming): "The Long Run Impact of Bombing Vietnam," Journal of Development Economics.

Prunier, G. (1998): The Rwanda Crisis: History of a Genocide. London: Hurst & Company.

Raleigh, C., A. Linke, H. Hegre, and J. Karisen (2010): "Introducing ACLED: An Armed Conflict Location and Event Dataset," Journal of Peace Research, 47(5), 651–660.

Restrepo, J., M. Spagat, and J. Vargas (2006): "The Severity of the Colombian Conflict: Cross-Country Datasets Versus New Micro-Data," Journal of Peace Research, 43(1), 99–115.

Reyntjens, F. (2009): The great African war: Congo and regional geopolitics, 1996-2006. New York, N.Y.: Cambridge University Press.

Shemyakina, O. (Forthcoming): "The Effect of Armed Conflict on Accumulation of Schooling: Results from Tajikistan," Journal of Development Economics, (12).

Timaeus, I. M. (1986): "Estimation of Mortality from Orphanhood in Adulthood," Demography, 28(2), 213–227.

Venieris, Y., and D. Gupta (1986): "Income distribution and socio-political instability as determinants of savings: A cross-sectional model," Journal of Political Economy, 96, 873–883.

Verpoorten, M. (2005): "The death toll of the Rwandan genocide: a detailed analysis for Gikongoro Province," Population, 60(4), 331–368.

——— (2009): "Cattle Sales in War and Peacetime: a Study of Household Coping in Rwanda, 1991-2001," Journal of Development Economics, 88, 67Ű86.

——— (2010a): "Detecting Hidden Violence: The Spatial Distribution of Excess Mortality in Rwanda," Discussion Paper 254, LICOS - Centre for Institutions and Economic Performance, K.U.Leuven.

——— (2010b): "The intensity of the Rwandan genocide: fine measures from the gacaca records," Discussion Paper 256, LICOS - Centre for Institutions and Economic Performance, K.U.Leuven.

Verpoorten, M., and L. Berlage (2007): "Economic Mobility in Rural Rwanda: A Study of the Effects of War and Genocide at the Household Level," Journal of African Economies, 16(3), 349–392.

Verwimp, P., and J. V. Bavel (2004): "Child Survival and the Fertility of Refugees in Rwanda after the Genocide," PRUS Working Papers 26, Poverty Research Unit at Sussex, University of Sussex.

Yanagizawa, D. (2010): "Propaganda and Conflict: Theory and Evidence from the Rwandan Genocide," Unpublished manuscript.

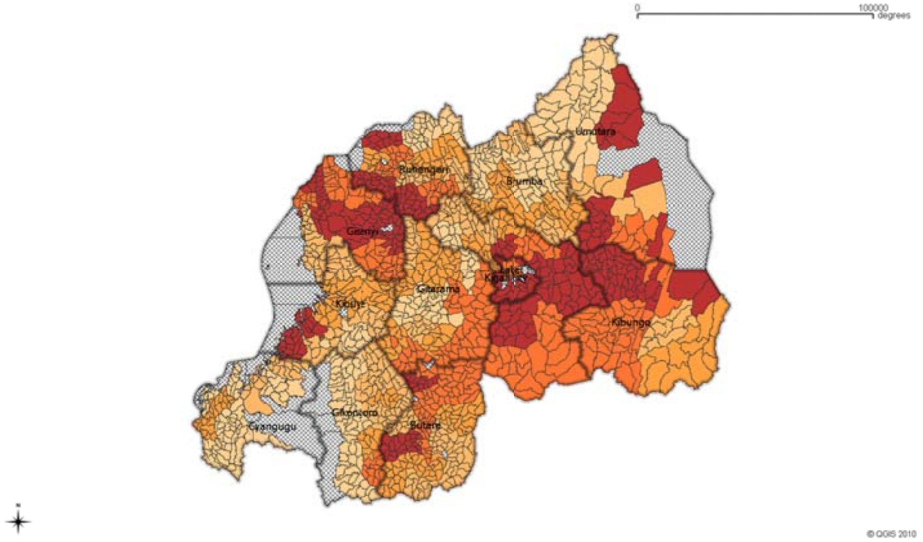
# Figures & Tables

**Figure 1. Geographical spread of genocide and civil war**

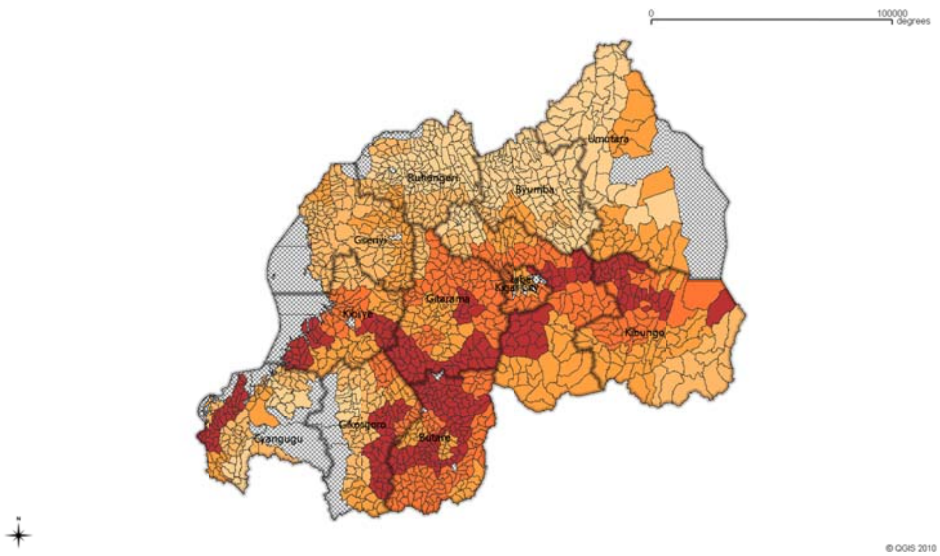

Notes: map taken from shape file of Rwandan provinces; location of different forms of violence based on event data; civil war includes the 1991-1994 civil war as well as the 1994-1998 (counter-)insurgencies between the RPF and the FAR.

**Figure 2: Map of respectively WEMI, GEMI and CEMI quintiles**

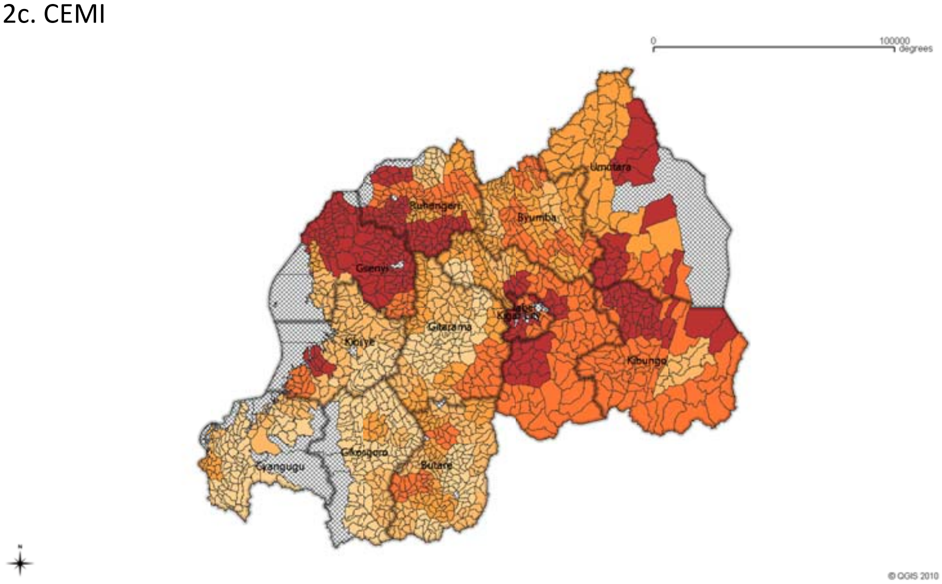
2a. WEMI




2b. GEMI

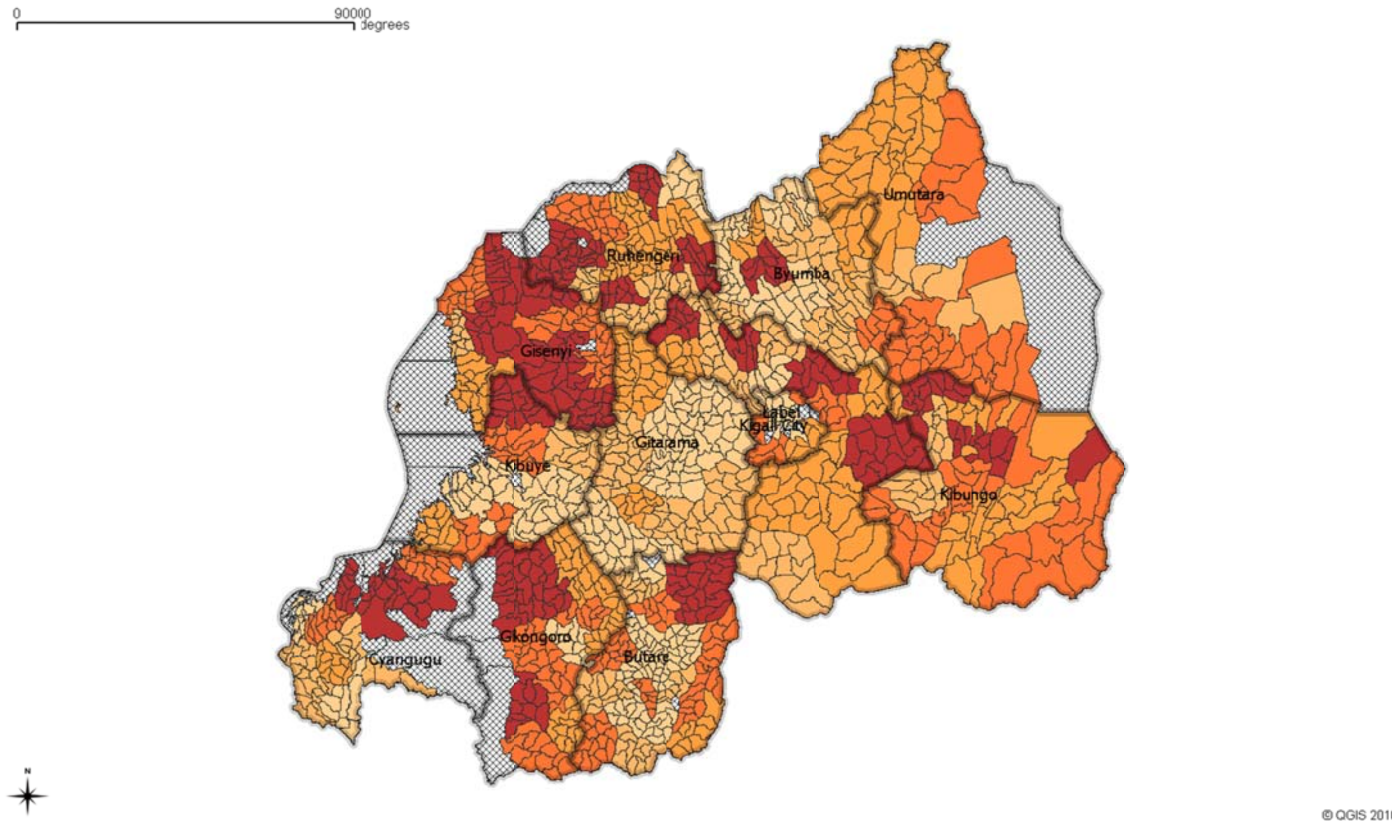



2c. CEMI




Notes: top quintile in darkest shade; map taken from shape file of Rwandan administrative sectors: quintiles calculated at commune level (which is one level above the sector, but for which no shape file exists)

**Figure 3. Commune level difference-in-difference estimates of the 1991-2002 schooling deficit**




Notes: top quintile (= largest schooling deficit) in darkest shade; map taken from shape file of Rwandan administrative sectors: quintiles calculated at commune level (which is one level above the sector, but for which no shape file exists)

**Table 1. Typology of the micro-empirical studies on the legacy of civil war**

| Type(s)* | Measurement of conflict exposure | Outcome studied | Country | Reference |
|---|---|---|---|---|
| I | Age cohorts | Education & Health | Cambodia | De Walque, 2004 |
| I, II | Age cohorts & Three different measures of province level conflict intensity measures (Nr=11) | Education | Rwanda | Akresh and De Walque, 2008 |
| I, II | Age cohorts & Distribution of the number of victims and human rights violations across 22 departments (Nr=22) | Education | Guatemala | Chamarbagwala and Morán, 2008 |
| I, II | Age cohorts & Rich district level data of the Peruvian Truth and Reconciliation Commission (CVR) (Nr=1833) | Education | Peru | Leon, 2010 |
| I, II | Age cohorts & Dummy for high conflict intensity provinces (Nr=11) | Height-for-age z-scores | Rwanda | Akresh et al., forthcoming |
| I, II | Age cohorts & Dummy for high conflict intensity provinces (Nr=17) | Height-for-age z-scores | Burundi | Bundervoet et al., 2009 |
| I, II | Age cohorts & Dummy for child being born in resettlement village (Nhh = 400) | Height-for-age z-scores | Zimbabwe | Alderman et al., 2006 |
| I, III | Age cohorts & Household damage dwelling reports (Nhh=1580) and distrcit level data on exposure to conflict (Nr=56) | Education | Tajikistan | Shemyakina, forthcoming |
| II | First PC of nr of assassinations, kidnappings, guerrilla attacks and displaced population). (Nr=55) | Household farm efficiency | Colombia | Gonzales and Lopez, 2005 |
| II | Bombing intensity at the district level (Nr=584) | Socio-economic outcomes | Vietnam | Miguel and Roland, 2010 |
| II | Railroad line damage (N=17 railroad lines) | Investment | China | Li, 2007 |
| II | Dummy for high conflict intensity provinces (Nr=17) | Household activity portfolio | Burundi | Bundervoet, 2007 |
| II | Dummy for high conflict intensity provinces (Nr=11) | Income growth | Rwanda | Justino and Verwimp, 2006 |
| II, III | The chiefdom average of four household conflict experience questions (Nr=153) | Local institutions | Sierra-Leone | Bellows and Miguel, 2009 |
| II, III | The community average of the incidence of civil strife, theft and physical attacks (Nr=370) | Non-agricultural enterprise start-ups | Uganda | Deininger, 2003 |
| III | Five household conflict experience questions (Nhh=256) | Income and asset mobility | Rwanda | Verpoorten & Berlage, 2007 |
| III | Five household conflict experience questions (Nhh=256) | Cattle sales | Rwanda | Verpoorten, 2009 |

*Types I : identification of impact using conflict exposure in time, II: identification of impact using conflict exposure in space, III : identification of impact using household conflict experience questions

(Nr=) Number of geographic entities; (Nhh=) Number of households

**Table 2. Commune level 1991 and 2002 excess mortality proxies (N=145)**

| Variable | Description | MP (1991 census) | | MP' (2002 census) | | WEMP (First difference) | |
|---|---|---|---|---|---|---|---|
| | | mean | st.dev. | mean | st.dev. | mean | st.dev. |
| Mortality of sons | Total number of boys died/number of boys born | 0.20 | 0.03 | 0.28 | 0.04 | 0.08 | 0.03 |
| Mortality of daughters | Total number of girls died/number of girls born | 0.18 | 0.03 | 0.24 | 0.03 | 0.07 | 0.03 |
| Widowhood | Widows among women (>=30 years) (% women >=30) | 0.18 | 0.03 | 0.31 | 0.05 | 0.12 | 0.05 |
| Double orphanhood | Double orphans among children and youngsters (<30years) (% children and youngsters <30) | 0.02 | 0.01 | 0.05 | 0.02 | 0.03 | 0.02 |
| Disability | Disabled due to war or genocide (% population) | 0.000 | 0.000 | 0.003 | 0.002 | 0.003 | 0.002 |

Notes: MP and MP' are vectors of respectively pre-war and post-war mortality proxies; WEMP are Wartime Excess Mortality Proxies", calculated as the first difference of MP' and MP.

**Table 3. Commune level genocide proxies (N=145)**

| | | Nationwide total | % 1994 population Mean | St. Dev. |
|---|---|---|---|---|
| (GP1) Category 1 suspects | Genocide suspects accused of planning, organizing or supervising the genocide, and committing sexual torture | 76,650 | 1.2% | 0.9% |
| (GP2) Category 2 suspects | Genocide suspects accused of killings or other serious physical assaults | 432,670 | 6.8% | 4.4% |
| (GP3) Category 3 suspects | Genocide suspects accused of looting or other offences against property | 309,500 | 5.0% | 3.5% |
| (GP4) Widowed genocide survivors | Widow or widower genocide survivors | 28,061 | 0.5% | 0.4% |
| (GP5) Orphaned genocide survivors | Orphaned genocide survivors | 75,078 | 1.2% | 1.0% |
| (GP6) Disabled genocide survivors | Disabled genocide survivors | 12,191 | 0.2% | 0.2% |

Notes: The proxies are taken from the records of the transitional justice courts (gacaca); the 1994 population is projected forward from the 1991 population census using commune level 1978-1991 population growth rates.

**Table 4. Educational attainment in 2002 and in 1991**

| | All Individuals aged 6-50 | | | |
|---|---|---|---|---|
| | 2002 | | 1991 | |
| Educational Attainment: | Obs. | % | Obs. | % |
| No Schooling | 174,363 | 30.67 | 203,459 | 39.36 |
| Some Primary Schooling | 347,676 | 61.15 | 286,972 | 55.52 |
| Some Secondary Schooling | 46,526 | 8.18 | 26,457 | 5.12 |
| Observations | 568,565 | 100 | 516,888 | 100 |

Notes: based on schooling attainment of individuals aged 6-50 in 10% random samples of the 1991 and 2002 population census

**Table 5. Difference-in-Differences comparing pre- and post-conflict schooling for young and old cohorts**

| Years of Schooling | Census 2002 | Census 1991 | Difference |
|---|---|---|---|
| Old Cohort | 3.727*** | 2.934*** | 0.793*** |
| | [0.007] | [0.007] | [0.010] |
| Young Cohort | 2.651*** | 2.853*** | -0.202*** |
| | [0.004] | [0.005] | [0.007] |
| Difference | -1.076*** | -0.081*** | -0.955*** |
| | [0.009] | [0.009] | [0.012] |

Notes: st.errors between brackets; young cohort = 6-22; old cohort = 23-50

**Table 6: Difference-in-Difference-in-Differences measuring the impact of armed conflict on years of schooling (OLS)**

| Dependent Variable: Years of Schooling | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| WEMI * (Young Cohort * CS 2002) | **-0.588*** ** | | | |
| | **[0.053]** | | | |
| WEMI*CS2002 | 0.311*** | | | |
| | [0.044] | | | |
| WEMI*Young cohort | -0.276*** | | | |
| | [0.038] | | | |
| WEMI | 0.258*** | | | |
| | [0.032] | | | |
| GEMI * (Young Cohort * CS2002) | | **-0.131*** ** | | **-0.113** ** |
| | | **[0.051]** | | **[0.051]** |
| GEMI*CS 2002 | | 0.073* | | 0.067 |
| | | [0.042] | | [0.042] |
| GEMI*Young cohort | | -0.360*** | | -0.361*** |
| | | [0.036] | | [0.036] |
| GEMI | | 0.438*** | | 0.492*** |
| | | [0.031] | | [0.031] |
| CEMI * (Young Cohort * CS2002) | | | **-0.629*** ** | **-0.609*** ** |
| | | | **[0.055]** | **[0.055]** |
| CEMI*CS 2002 | | | 0.346*** | 0.319*** |
| | | | [0.046] | [0.045] |
| CEMI*Young cohort | | | -0.150*** | -0.160*** |
| | | | [0.039] | [0.039] |
| CEMI | | | 0.069** | -0.039 |
| | | | [0.034] | [0.035] |
| Young Cohort * CS 2002 | -0.922*** | -1.128*** | -0.922*** | -0.899*** |
| | [0.024] | [0.021] | [0.023] | [0.029] |
| CS 2002 | 0.528*** | 0.637*** | 0.521*** | 0.515*** |
| | [0.020] | [0.018] | [0.019] | [0.024] |
| Household Level Controls? | Yes | Yes | Yes | Yes |
| Child Age Fixed Effects? | Yes | Yes | Yes | Yes |
| Province Fixed Effects? | Yes | Yes | Yes | Yes |
| Observations | 1,074,561 | 1,074,561 | 1,074,561 | 1,074,561 |

Note: st. errors between brackets; the control variables include indicator variables for being female, being non-poor and living in a rural area, the age of the household head, the highest years of education of any adult household member and the number of children aged 5 or less.

**Table 7: Difference-in-Difference-in-Differences measuring the impact of armed conflict on years of schooling (IV 2nd stage)**

| Dependent Variable: Years of Schooling | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| WEMI * (Young Cohort * CS 2002) | **-0.594*** ** | | | |
| | **[0.051]** | | | |
| WEMI*CS2002 | 0.367*** | | | |
| | [0.042] | | | |
| WEMI*Young cohort | -0.635*** | | | |
| | [0.037] | | | |
| WEMI | 0.870*** | | | |
| | [0.039] | | | |
| GEMI * (Young Cohort * CS2002) | | **0.012** | | **-0.145*** ** |
| | | **[0.053]** | | **[0.055]** |
| GEMI*CS 2002 | | 0.008 | | 0.096** |
| | | [0.044] | | [0.046] |
| GEMI*Young cohort | | -0.727*** | | -0.872*** |
| | | [0.000] | | [0.040] |
| GEMI | | 1.700*** | | 2.222*** |
| | | [0.041] | | [0.046] |
| CEMI * (Young Cohort * CS2002) | | | **-0.486*** ** | **-0.506*** ** |
| | | | **[0.041]** | **[0.042]** |
| CEMI*CS 2002 | | | 0.301*** | 0.299*** |
| | | | [0.034] | [0.035] |
| CEMI*Young cohort | | | -0.267*** | -0.437*** |
| | | | [0.030] | [0.030] |
| CEMI | | | 0.030 | -0.713*** |
| | | | [0.040] | [0.044] |
| Young Cohort * CS 2002 | -0.801*** | -1.142*** | -0.924*** | -0.808*** |
| | [0.029] | [0.029] | [0.020] | [0.038] |
| CS 2002 | 0.408*** | 0.551*** | 0.514*** | 0.341*** |
| | [0.024] | [0.024] | [0.017] | [0.031] |
| Household Level Controls? | Yes | Yes | Yes | Yes |
| Child Age Fixed Effects? | Yes | Yes | Yes | Yes |
| Province Fixed Effects? | Yes | Yes | Yes | Yes |
| Observations | 1,074,561 | 1,074,561 | 1,074,561 | 1,074,561 |

Note: st. errors between brackets; same control variables as in Table 6.

**Table 8: First stage of IVE**

| Dependent Variable: | WEMI | GEMI | CEMI |
|---|---|---|---|
| *Excluded instruments:* | | | |
| log(distance to Nyanza) | -0.060*** | -0.082** | -0.018 |
| | [0.039] | [0.037] | [0.034] |
| log(distance to Uganda) | 0.084*** | 0.052* | 0.067** |
| | [0.030] | [0.028] | [0.026] |
| *Included instruments:* | | | |
| Household Level Controls | Yes | Yes | Yes |
| Province Fixed Effects? | Yes | Yes | Yes |
| Observations | 145 | 145 | 145 |

Note: st. errors between brackets; same control variables as in Table 6 but no age fixed effects and household level controls averaged at the community level.

**Table 9: Difference-in-Difference-in-Differences measuring the impact of armed conflict on years of schooling, after correcting the conflict indices for survival bias**

| Dependent Variable: Years of Schooling | OLS | | IV 2nd stage | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| WEMI * (Young Cohort * CS 2002) | **-0.595*** ** | | **-0.528*** ** | |
| | **[0.052]** | | **[0.041]** | |
| GEMI * (Young Cohort * CS2002) | | **-0.153*** ** | | **-0.165*** ** |
| | | **[0.049]** | | **[0.055]** |
| CEMI * (Young Cohort * CS2002) | | **-0.581*** ** | | **-0.509*** ** |
| | | **[0.054]** | | **[0.038]** |
| All constitutive terms of the interaction effect? | Yes | Yes | Yes | Yes |
| Household Level Controls? | Yes | Yes | Yes | Yes |
| Child Age Fixed Effects? | Yes | Yes | Yes | Yes |
| Province Fixed Effects? | Yes | Yes | Yes | Yes |
| Observations | 1,074,561 | 1,074,561 | 1,074,561 | 1,074,561 |

Note: st.errors between brackets; same control variables as in Table 6.

**Table 10: Difference-in-Difference-in-Differences results on the impact of armed conflict on years of schooling across alternative measures of violent conflict intensity (OLS and IV 2nd stage)**

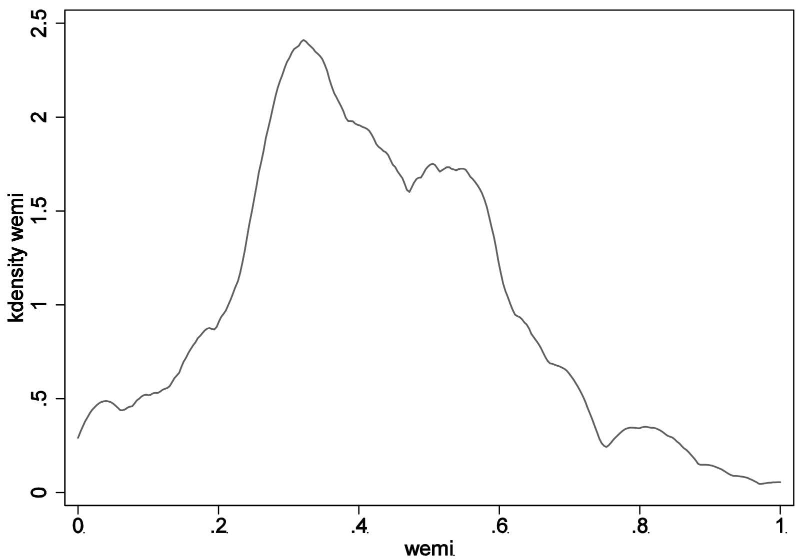
| Dependent Variable: Years of Schooling | | Difference-in-Difference-in-Difference estimate | | | |
|---|---|---|---|---|---|
| Conflict intensity measure used | | OLS | | IV | |
| ***Province level measures*** | | | | | |
| *Event data* | | | | | |
| **C1** | '94 genocide dummy | 0.168*** | [0.021] | 0.264*** | [0.050] |
| **C2** | '90-'92 civil war dummy | 0.042* | [0.027] | -0.066*** | [0.047] |
| **C3** | '90-'94 civil war dummy | -0.135*** | [0.019] | -0.363*** | [0.052] |
| **C4** | '94-'98 (Counter-)insurgency dummy | -0.119*** | [0.024] | -0.005 | [0.052] |
| *Genodynamics project* | | | | | |
| **C5** | '94 death toll dummy | -0.139*** | [0.021] | 0.194*** | [0.054] |
| **C6** | Days with killings in April-July '94 | 0.155*** | [0.046] | 0.160*** | [0.054] |
| *ACLED* | | | | | |
| **C7** | Number of battle events | 0.023 | [0.036] | 0.094* | [0.087] |
| *1991 Population census* | | | | | |
| **C8** | Proportion of Tutsi | 0.029 | [0.031] | -0.033 | [0.058] |
| *Yale genocide studies* | | | | | |
| **C9** | Number of mass graves | 0.006 | [0.028] | 0.131*** | [0.050] |
| **C10** | Distance to mass grave | 0.037 | [0.036] | -0.011 | [0.055] |
| *Gacaca records* | | | | | |
| **C11** | Proportion of genocide suspects | 0.077*** | [0.027] | 0.240*** | [0.048] |
| ***Commune level measures*** | | | | | |
| *1991 Population census* | | | | | |
| **C12** | Proportion of Tutsi | -0.165** | [0.078] | -0.064 | [0.045] |
| *Yale genocide studies* | | | | | |
| **C13** | Number of mass graves | -0.357*** | [0.050] | -0.143*** | [0.043] |
| **C14** | Distance to mass grave | 0.097* | [0.055] | 0.098** | [0.050] |
| *Gacaca records* | | | | | |
| **C15** | Proportion of genocide suspects | 0.018 | [0.042] | 0.223*** | [0.037] |
| Observations | | 1,074,561 | | 1,074,561 | |

Note: st.errors between brackets; province and age fixed effects; same control variables as in Table 6.

# Appendix

**Figure A1: kernel density of the conflict indices**

a. WEMI




b. GEMI and CEMI

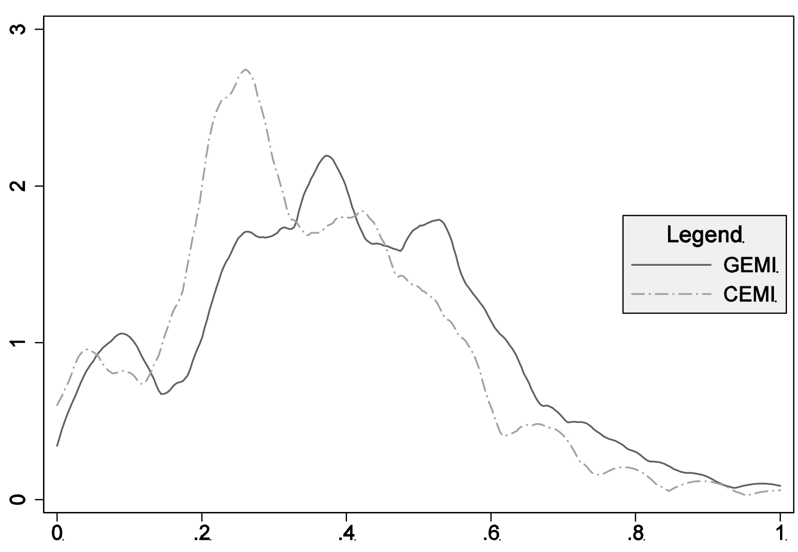


**Table A1: Overview of conflict intensity measures for Rwanda**

| | | Used in | Mean | St. dev. | Correlation with | | |
|---|---|---|---|---|---|---|---|
| ***Province level measures*** | | | | | WEMI | GEMI | CEMI |
| *Event data* | | | | | | | |
| **C1** | Genocide dummy | Justino & Verwimp, 2006 | 0.35 | 0.48 | -0.32*** | 0.31*** | -0.54*** |
| **C2** | '90-'92 civil war dummy | Akresh et al., forthcoming | 0.25 | 0.43 | -0.09*** | -0.59*** | 0.24*** |
| **C3** | '90-'94 civil war dummy | Justino & Verwimp, 2006 | 0.44 | 0.50 | 0.17*** | -0.38*** | 0.42*** |
| **C4** | '94-'98 (Counter-)insurgency dummy | Justino & Verwimp, 2006 | 0.20 | 0.40 | 0.27*** | -0.379*** | 0.51*** |
| *Genodynamics project* | | | | | | | |
| **C5** | '94 death toll dummy | Akresh & de Walque, 2008 | 0.32 | 0.46 | 0.28*** | 0.41*** | 0.09*** |
| **C6** | Days with killings in April-July '94 | Akresh & de Walque, 2008 | 24.22 | 20.39 | 0 | 0.31*** | -0.18*** |
| *ACLED* | | | | | | | |
| **C7** | Number of battle events | N.A. | 30.87 | 18.11 | 0.07 | -0.38*** | 0.25*** |
| *1991 Population census* | | | | | | | |
| **C8** | Proportion of Tutsi | Akresh & de Walque, 2008 | 0.08 | 0.06 | -0.11*** | 0.58*** | -0.46*** |
| *Yale genocide studies* | | | | | | | |
| **C9** | Number of mass graves | Akresh & de Walque, 2008 | 6.28 | 4.88 | 0.10*** | 0.56*** | -0.21*** |
| **C10** | Distance to mass grave | N.A. | 12.29 | 7.81 | -0.13*** | -0.51*** | 0.15*** |
| *Gacaca records* | | | | | | | |
| **C11** | Number of perpetrators | N.A. | 0.07 | 0.04 | -0.05*** | 0.66*** | -0.43*** |
| ***Commune level measures*** | | | | | | | |
| *1991 Population census* | | | | | | | |
| **C12** | Proportion of Tutsi | N.A. | 0.12 | 0.14 | 0.17*** | 0.68*** | -0.17*** |
| *Yale genocide studies* | | | | | | | |
| **C13** | Mass grave | N.A. | 0.47 | 0.81 | 0.20*** | 0.47*** | -0.03 |
| **C14** | Distance to mass grave | N.A. | 12.29 | 10.17 | -0.22*** | -0.53*** | 0.05* |
| *Gacaca records* | | | | | | | |
| **C15** | Number of perpetrators | Yanagizawa, mimeo | 0.07 | 0.05 | 0.14*** | 0.85*** | -0.33*** |

## Description of 15 alternative conflict intensity measures

This appendix describes 15 conflict intensity measures for Rwanda (denoted C1-C15). These measures are used to produce the results in Table 10. Table A1 gives an overview.

Eleven (C1-C11) of the 15 measures are at the province level, the remaining four (C12-C15) are at the commune level. The measures rely on six distinct primary data sources.

1. Event data

- C1: '94 genocide dummy
- C2: '90-'92 civil war dummy
- C3: '90-'94 civil war dummy
- C4: '94-'98 (Counter-)insurgency dummy

Justino and Verwimp (2006) rely on event data[1] to evaluate the intensity of genocide, civil war and counter-insurgency in Rwanda across provinces. They identify four provinces with high 1994 genocide intensity (Butare, Cyangugu, Kibuye, Gikongoro - C1). The provinces are further categorized in provinces with high (Kibungo, Rural Kigali, Ruhengeri and Byumba - C3) and low 1990-1994 civil war intensity and provinces with high (Ruhengeri and Gisenyi - C4) and low levels of 1994-1998 (counter)-insurgency. In a study on the impact of the 1990-1992 civil war on height-for-age z-scores, Akresh et al. (forthcoming) use an indicator variable for provinces exposed to civil war in 1990-1992 (Byumba and Ruhengeri - C2).

2. Genodynamics project

- C5: '94 death toll dummy
- C6: Days with killings in April-July '94

In their study on the impact of armed conflict on schooling, Akresh & de Walque (2008) use a conflict intensity dummy that takes one for Butare, Rural Kigali, and Kibungo, the three provinces with the highest 1994 death toll. In addition, they use the province level number of days with killings in the period April-June 1994. Both measures are taken from the Genodynamics project[2], which compiled data on 1994 casualties from different sources, including The Ministry of Education in Rwanda, Ministry of Youth, Culture and Sports in Rwanda, IBUKA (an association of Tutsi survivors), African Rights and Human Rights Watch (both international human rights organizations).

3. ACLED

- C7: Number of battle events

The Armed Conflict Location and Event Data (ACLED), described in Raleigh et al. (2010), record 322 battle events in Rwanda for the period 1990-2002. The precision of the geographical information varies across battle events, with about two thirds identified at the commune level and one third at the province level. Therefore, the information is aggregated at the province level.

4. 1991 population census

- C8 & C12: Province and commune level proportion of Tutsi in the population

---

[1] e.g. the Human Rights Watch report of Des Forges (1999) which is largely based on witness accounts.

[2] For more information on the Genodynamics project, see Davenport and Stam (2009) and http://www.genodynamics.com.

From the 1991 population census, one can derive the proportion of Tutsi both at the province level (C8) and at the commune level (C12). This is useful information because Tutsi were targeted in the genocide and their proportion in the population varies widely across as well as within provinces.

5. Yale Genocide Studies

- C9 & C13 : Province and commune level number of mass graves
- C10 & C14: Province and commune level distance to nearest mass grave

On the website of Yale Genocide Studies, one can find a map with the location of 71 mass graves in Rwanda. The map was realized by the Rwandan commission of genocide memorial. The number of mass grave sites and memorials per province is used as a third measure of war intensity in the paper of Akresh & De Walque (2008). We look at two related measures, which are defined at the province level (C9-C10) as well as the commune level (C13-C14): the number of massgraves and the distance to the nearest mass grave. The distance to the nearest mass grave (in km) is calculated by overlaying a geo-referenced administrative map with the location of 71 mass graves in Rwanda.

6. Gacaca records

- C11&C15 Province and commune level proportion of genocide suspects

Over the period 2005-2007, the transitional Rwandan justice system in charge of judging 1994 genocide suspects, referred to as gacaca, collected information on genocide victims, suspects and survivors (Government of Rwanda, 2005). The sector level numbers of genocide suspects and survivors were made public in the course of 2007. In a study of the impact of propaganda (through radio transmission) on civilian participation in the genocide, Yanagizawa (2010) makes use of the commune level proportion of genocide suspects to proxy participation (C11&C15).