Measure for Measure:

How Well Do We Measure Micro-level Conflict Intensity?

Marijke Verpoorten*

September, 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: a fine Wartime Excess Mortality Index (WEMI). It is argued that the proposed measure is particularly well suited for studying the legacy of civil wars that are characterized by a large death toll and by different forms 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, the estimated impact varies widely across WEMI and a large set of alternative conflict intensity measures for Rwanda. While the conflict intensity measure proposed in this paper requires further study and one probably needs a combination of various methodologies, this finding suggests the need for a careful understanding of what underlies the different measures and methodologies in use.

JEL: C81, 015, C21

Armed Conflict, Micro-level Conflict Measures, Rwanda, Schooling

*Research scholar of the Fund for Scientific Research - Flanders, Research associate of LICOS - KULeuven, Assistant Professor at IOB - University of Antwerp. 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. For example, as argued in a recent literature overview by Blattman & 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". New data is key to successfully embarking on this challenge. Again, in the words of Blattman & Miguel: “A major goal of civil war researchers should be the collection of new data”. Or more specifically, in the words of Restrepo et al. (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 first reviews commonly used measures of conflict intensity and then demonstrates how existing and widely available data can form the basis of a useful subnational measure of conflict intensity.

The newly proposed measure is not meant to replace existing measures, but is meant to be complementary and is particularly useful for measuring multidimensional conflict characterized by 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 [1]. Moreover, among the 12 countries with a post-conflict population census, 10 have had a pre-conflict population census, which provides useful information for calculating a geographically disaggregated mortality baseline. The census questionnaires include a set of comparable questions, informing about the respondent's sex, age, marital status, location and duration of previous and current residence, survival of parents and survival of children. Whereas census definitions of citizenship may be highly politicized (since citizenship defines who may vote), the questions that can be used to assess excess mortality (marital status, survival of children and survival of parents) are not.

Rwanda is taken as an illustration. This country in Africa's turbulent Great Lakes region experienced several forms of violence in the 1990s, including genocide, civil war, reprisal killings and (counter)insurgency (i.e. rural guerilla warfare). In order to measure Rwanda's multidimensional conflict cycle, I develop a spatial index of wartime excess mortality relying

[1] The exception is Somalia. UCDP/PRIO conflict data and UN population census statistics (http://unstats.un.org/unsd/demographic/sources/census/censusdates.htm)

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")[2].

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 countries, the proposed measure can be applied to other countries that have experienced armed conflict. 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. The latter is important since it is estimated that the number of indirect deaths of conflict is six times larger than the number of battle-related direct deaths (Human Security Report, 2010).

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 potential drawback is that other events unrelated to conflict may explain wartime excess mortality (e.g. a local harvest failure or 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.

[2]In the 1994 administrative subdivision of Rwanda, "prefectures" were followed by "communes", "sectors" and "cells". In a series of subsequent reforms in 1996 and 2002, prefectures and communes were replaced by provinces and districts, respectively .

The use of $WEMI$ 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 (DDD) estimation of years of education of a young and an older age cohort in the 1991 and 2002 population census, with the treated group being the young age cohort in the 2002 census residing in high conflict intensity regions. I find that both the OLS and the IVE results strongly depend on the type of conflict intensity measure used.

The next section gives an overview of the conflict intensity measures used in the micro-empirical literature on the legacy of violent conflict. Then, I outline the method for the newly proposed measure, and illustrate it for the Rwandan conflict cycle. Finally, I provide an application studying the impact of violent conflict on schooling in Rwanda, and compare the results across different conflict intensity measures.

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]. 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. Each type is evaluated on its strengths and weaknesses. Table I summarizes the studies by type.

Insert Table I about here

Type I: conflict exposure in time

The first set of studies measure conflict exposure in time, 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 (DD) estimates relying on two nationwide

[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 et al., 2002).

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 DDD estimation, relying on variation in conflict intensity across time as well as space.

Examples of type-I studies are Akresh & de Walque (2008), Akresh et al. (2011), Alderman et al. (2006), Bundervoet et al. (2009), Chamarbagwala & Moran (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 I 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, regions or other administrative units 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 et al., 2011; Bundervoet et al., 2009) as well as Bundervoet (2007) and Justino & 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 & Moran, 2008; and Akresh & de Walque, 2008), as well as an additional three studies: Gonzales & Lopez (2007), Miguel & Roland (2011) and Li (2007). Gonzales & Lopez (2007), looking at the effect of political violence in Columbia on farm household efficiency, subject detailed data on human rights violations to principal component analysis. The variables taken into account are 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 at the municipal level (Nr=55). Miguel and Roland (2011) 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 often 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 et al., 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, which on average yields good results. However, in some cases press accounts may be biased and computer news screening may be sensitive to the language in which events are reported. Another drawback, highlighted by Restrepo et al. (2006), is that battle events (e.g. between a rebel group and government troops) may come short in reflecting violence against 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 & Miguel, 2009; Deininger, 2003; Shemyakina, 2011). This set of studies overlaps with the third type.

Type III: Household conflict experience questions

Studies of type III rely on household surveys. A distinction can be made between standard surveys (e.g. the Integrated Household Living Conditions Survey, IHLCS) and surveys that include a special module on household conflict experience. The standard surveys often include questions on migration status, damage to household dwellings and asset loss, and these questions can be used as proxies for refugee experience and wartime shocks to physical capital. Studies using such information include those that analyze socioeconomic outcomes across refugee and non-refugee households (e.g. Kondylis, 2008; Verwimp & Van Bavel, 2004), as well as one previously mentioned type-I study, Shemyakina (2011), which uses information on damage to household dwellings to study the impact of armed conflict in Tajikistan on schooling outcomes.

The limited number of surveys including a special module on conflict experience inquire about direct confrontations with violence, e.g. as a perpetrator or as a victim. These surveys are often the result of the researchers' own 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 & Miguel (2009), Deininger (2003), Verpoorten & 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 uniform measure of conflict intensity has triggered of a debate on the inclusion of a conflict module in standard household surveys (Brück et al., 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. One potential drawback is that, 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 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.

A new measure of conflict intensity

This section outlines the method for a new spatial measure of conflict exposure, which is constructed in three steps from two rounds of population census data. 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[4]. These proxies may include widowhood (%), orphanhood (%) and mortality of sons and daughters (% of life births). The pre- and post-war vectors of p mortality proxies at the level of administrative unit j can be written as follows:

$$\begin{aligned} MP_j^{pre-war} &= [MP_{j1}, MP_{j2}, ..., MP_{jp}], \\ MP_j^{post-war} &= [MP'_{j1}, MP'_{j2}, ..., MP'_{jp}]. \end{aligned} \tag{1}$$

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 ($WEMP_j$):

$$MP_j^{post-war} - MP_j^{pre-war} = WEMP_j = [WEMP_{j1}, WEMP_{j2}, ..., WEMP_{jp}]. \tag{2}$$

Third, the set of wartime excess mortality proxies is summarized into an index by taking a weighted sum, with the weights determined by PCA, which has the desirable property of reducing the dimensionality of a variable set while retaining as much as possible of the variation present [5]. The first PC will be an appropriate summary of excess mortality on a

[4]Provided that this smallest common administrative unit is large enough in terms of population to reduce the impact of outliers and erroneous data.

[5]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 α_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 α_1 maximizes $var[\alpha_1' x^*] = \alpha_1' \Sigma \alpha_1$, with Σ 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 α_1 as the eigenvector of Σ corresponding to the largest eigenvalue of Σ, λ_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 α_{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 α_{1l}/s_l.For a detailed exposition of principal component analysis I refer to Jolliffe (2002).

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 (Jolliffe, 2002). The first PC, referred to as the Wartime Excess Mortality Index, can be written as:

$$WEMI_j = l \times WEMP_j = l \times [WEMP_{j1}, WEMP_{j2}, ..., WEMP_{jp}] \quad (3)$$

with l the vector of loadings obtained through PCA.

A number of studies have used PCA for the purpose of aggregating conflict indicators. 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 & Gupta, 1986; Barro, 1991; Alesina & Perotti, 1996). To the best of our knowledge, only one micro-economic study, Gonzales & Lopez (2007) - discussed in the previous section, uses PCA to summarize variables into a micro level index of violent conflict. 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 & Trussel, 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 within a given country on a less to more scale.

Illustration: measuring the Rwandan conflict cycle

The Rwandan conflict cycle

The Rwandan conflict cycle of the nineties included civil war, genocide, reprisal killings, (counter)insurgency and a major refugee crisis. While these events all occurred in the nineties, their geographic location within Rwanda 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 1995-1998 (counter)insurgency concentrated in the Northwest (Des Forges, 1999; Davenport & Stam, 2009). This broad spatial pattern is illustrated in Figure I.

Insert Fig. 1 about here

Among the events in the Rwandan conflict cycle, the genocide had by far the largest direct death toll, with an estimated 800,000 Tutsi and moderate Hutu killed in a time span of barely 100 days (Des Forges, 1999; Prunier, 1998; Verpoorten, 2005). There is much less accurate information on the death toll of the civil war, but it is likely that tens of thousands of people became victims of the fighting between the RPF and the FAR, or fell victim to reprisal killings by the RPF (Davenport & Stam, 2009; Reyntjens, 2009; the "Gersony report"[6]). Regarding the two latter forms of violence, Davenport & Stam (2009) estimate that, during April-June 1994, the number of individuals killed in zones under RPF control and the zones contested by the RPF and FAR amounted to respectively 80,000 and 90,000.

There is even more uncertainty about the number of indirect deaths of the conflict. Many may have died prematurely following the collapse of health care, social and economic systems. Furthermore, the death toll in refugee camps was very high due to the rapid spread of infectious diseases. For example, the cholera epidemic in Goma (at the border of RD Congo and Rwanda) is believed to have taken around 30,000 lives (Prunier, 1998). The indirect death toll in Gisenyi and Ruhengeri may have been higher than elsewhere because these northwestern provinces did not only serve as the corridor of approximately one million refugees fleeing to Congo in 1994 and back to Rwanda in 1996/1997, but they also experienced a relatively long period of violence as well as important disruptions in economic activities. In this respect, Amnesty International reports that, as part of the counterinsurgency strategy, a scorched earth policy was being carried out in many areas in the Northwest, where homes and fields were being burned. In addition, they report that, in an attempt to cut food supplies to armed opposition groups, the RPF prevented farmers from harvesting and marketing their crops (Amnesty International, 1997).

[6]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).

The Wartime Excess Mortality Index for Rwanda

For the construction of $WEMI$ for the Rwandan conflict, 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)[7]. The 2002 census includes information at the level of the administrative sectors (N=1540), while 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 of 174 squared km and approximately 55,000 inhabitants on average.

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);
- ($WEMP_{j3}$) ΔWidowhood : 2002-1991 difference in the proportion of widows (among women $>$ 30 years);
- ($WEMP_{j4}$) ΔDouble 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 (only applicable in the 2002 census).

For the calculation of widowhood, I set the lower age limit at 30 because it is likely that a considerable share of women aged below 30 in 2002 were not yet married in 1990, the start of the conflict, and those who were married may have remarried upon widowhood, given their young age. For orphans, usually one considers the age groups 0-15 or 0-18. Here, I take the age group 0-30 as a baseline since youngsters aged 18 at the start of the conflict were 30 years old by 2002. Below, I perform a sensitivity analysis with respect to the age limits.

[7] It has been demonstrated that the Rwandan census data are very reliable except for the recording of ethnicity. Verpoorten (2005) compares the 1991 Rwandan census data with the 1990 population data from the local administration and finds almost an exact match of numbers of women and men. However, the share of Tutsi in the population was underreported in the national census data, either by the Habyarimana regime to keep their school enrolment and public employment quotas low, either by Tutsi themselves to avoid discimination.

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. To gauge the impact of displacement on the results, I calculate the 2002 mortality proxies both including and excluding individuals who changed residence over the period 1990-2002. The latter can be derived from the 2002 census by combining individual-level information of current residence, previous residence and duration of current residence. Since the purpose is to retrieve a spatial pattern of wartime excess mortality, the baseline results are calculated excluding individuals who changed residence over the period 1990-2002, which leaves a sample of 7.4 million individuals in the 2002 census (91%). The results including migrants are basically similar (see below).

Table II provides summary statistics for the pre-war and post-war MPs and their first differences. All MPs have higher values in the 2002 population census than in the 1991 population census. The mortality of sons and daughters increased from 20% to 28% and from 18% to 24%, respectively. 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:

$$WEMI_j = 0.48 \times WEMP_{j1} + 0.43 \times WEMP_{j2} + 0.46 \times WEMP_{j3} \tag{4}$$

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

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 II plots the quintiles of $WEMI_j$ on an administrative map of Rwanda. Many top quintile communes are located in Butare Province, in and around Kigali City, in Gisenyi, in the northwestern corner of Kibungo and in the southwestern corner of Umutara. Smaller local clusters can be found in the West of Ruhengeri and Kibuye and in the Southeast of Gitarama and Gikongoro. Taken together, this map reflects well the spatial pattern of different forms of violence that took place in Rwanda (see Figure I). Moreover, elsewhere, it is demonstrated that the spatial pattern of $WEMI$ is strongly related to geographical determinants of violence in Rwanda, such as the commune-level share of Tutsi, the number of days that a commune was under RPF control in 1994, the distance to a main road, the distance to a refugee camp (from where the insurgents operated), and alleged human rights violations during the 1995-1998 (counter)insurgency. This is extensively discussed in Verpoorten (2010), who studies the determinants of the spatial pattern of excess mortality in Rwanda.

Insert Table II about here

Insert Figure II about here

Robustness tests

I perform several robustness tests to check the sensitivity of $WEMI$ to different specifications and to possible sources of bias. First, I calculate $WEMI$ based on the full census, i.e. including migrants. Second, I leave out the different wartime excess mortality proxies case by case to assess whether the first component unintentionally overemphasizes some mortality over others. Third, instead of including double orphanhood as a mortality proxy, I include maternal and paternal orphanhood. Fourth, I perform 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)).

Finally, I make a correction for possible survival bias. $WEMI$ 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 $WEMI$ is given by the following linear combination:

$$WEMI\_s_j = 0.47 \times WEMP_{j1} + 0.42 \times WEMP_{j2} + 0.45 \times WEMP_{j3}$$

$$+0.48 \times WEMP_{j4} + 0.38 \times WEMP_{j5} - 0.15 \times s_j,$$

with s_j "log(distance to mass grave)".

Table AI in Appendix gives an overview of these alternative calculations and lists their correlation coefficients with the baseline $WEMI$. All correlation coefficients are very close to 1, indicating that the spatial pattern of excess mortality is very robust to how exactly $WEMI$ is being calculated. Below I show that these measures also yield basically the same results in an empirical application.

Placebo tests

A crucial assumption made is that the variation in $WEMI$ mainly stems from conflict, and that other events such as local crop failures and related deaths or increases in HIV/AIDS deaths only play a relatively minor role. This assumption cannot be directly verified. However, indirect support for this assumption can be provided by comparing $WEMI$ in a context of armed conflict with $WEMI$ in a context in which no conflict took place. As a case, I use Tanzania, which is the only neighboring country of Rwanda that did not experience civil war in the 1990s. Possible sources of excess mortality in Tanzania include the HIV/AIDS epidemic, local harvest failures and spillover from the Rwandan conflict in the form of large refugee streams, but these are very unlikely to have resulted in excess mortality of the same magnitude as in Rwanda.

I rely on a 10% random draw of the 1988 and 2002 Tanzanian population census (N=2.3 million individuals and N= 3.7 million, resp.; accessed online from IPUMS International), and calculate a set of wartime excess mortality proxies at the level of 113 Tanzanian districts. The excess mortality proxies include the mortality of sons, the mortality of daughters and widowhood (defined as in MP_1, MP_2 and MP_3 above), and maternal orphanhood (instead of double orphanhood - MP_4 above -since the Tanzanian 1988 census does not provide information on the survival of the father. The results are given in Table AII in Appendix. It can be seen that - in contrast to the strong increases for MP_1, MP_2, MP_3 and maternal orphanhood in Rwanda - in Tanzania, MP_1 and MP_2 decrease over time, while MP_3 remains stable and MP_4 increases only slightly. Hence, this placebo test provides support for the assumption that the variation in the Rwandan WEMI is largely due to armed conflict.

Application: armed conflict and schooling in Rwanda

The aim of this section is to compare the results of an empirical analysis across different conflict measures, including $WEMI$. For this purpose, the application focuses on a well-established impact of armed conflict, i.e. its negative impact on schooling (e.g. Lai, 2007). The section proceeds in four steps. First, I present the data on schooling. Second, I calculate the schooling deficit over time in each of the 145 communes, obtaining a spatial pattern of the 1991-2002 schooling deficit. Third, I combine information on conflict exposure in time with the spatial information embodied in $WEMI$ to calculate DDD estimates of the schooling deficit. Finally, I compare these DDD results with those obtained using a large number of alternative conflict intensity measures for the Rwandan conflict.

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[8]. 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 in 2002 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 qualitatively similar results - not reported)[9].

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 III provides summary statistics on the number of years of schooling completed for the young and the old age cohort across 1991 and 2002. The figures show a progress of 0.8 years of schooling for the old age cohort and a drop of 0.2 years for the young age cohort, yielding a DD estimate of one year of schooling, or, when taking logged years of schooling as a dependent variable, a 19.3% decrease in the number of years of schooling.

This result is in line with the study of Akresh & 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. The present study uses census data instead of DHS data because the former have a complete geographic coverage and can be combined with commune level measures of conflict intensity (instead of province level measures as is the case for the DHS data).

Insert Table III about here

[8] I use only 10% of the 2002 census, because the regression analysis - run in STATA - cannot be executed using the entire census (8.1 million observations).

[9] The alternative age categories used include 8-17 and 6-22 for the young cohort and 18-37 and 23-40 for the old cohort.

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 and identify the spatial pattern of the schooling deficit. The commune level DD are obtained by estimating the following equation for each commune j separately:

$$Y_{ijt} = \alpha_{j0} + \alpha_{j1}(T_t \times young_i) + \alpha_{j2}T_t + \alpha_{j3}young_i + \varepsilon_{ijt} \quad (5)$$

WITH

Y_{ijt} : average years of schooling 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

ε_{ijt} : idiosyncratic error

The coefficients α_{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 III 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. Since the share of Tutsi in the northwestern provinces was as low as 1.5% (compared to over 10% in the South), the estimated schooling deficit in the Northwest cannot be attributed to the genocide, but is likely due to other events in the conflict cycle, primarily the 1995-1998 (counter)insurgency and the refugee crisis. Note that this finding stands in contrast with the study of Akresh & de Walque (2008) who implicitly attribute the entire estimated schooling deficit to genocide[10].

Insert Fig. 3 about here

[10] In an online appendix to this article, I show this more formally. I augment the set of mortality proxies subjected to PCA with a set of genocide-specific proxies, obtaining two indices, one for genocide-related excess mortality (GEMI) and one for other sources of excess mortality (CEMI). It is shown that GEMI can account for less than than half of the schooling deficit.

Difference-in-Difference-in-Difference estimates: explaining the 1991-2002 schooling deficit (OLS and IVE)

To determine causality, I take the analysis two steps further. First, I estimate a DDD model 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, I instrument for conflict intensity.

The DDD estimate corresponds to the coefficient β_1 in the following equation:

$$\begin{aligned} Y_{ijt} &= \beta_0 + \beta_1(T_t \times C_j \times young_i) + \beta_2(T_t \times C_j) + \beta_3(T_t \times young_i) \\ &\quad + \beta_4(C_j \times young_i) + \beta_5 T_t + \beta_6 C_j + \beta_7 X_{ijt} + \zeta_i + \eta_j + \varepsilon'_{ijt} \end{aligned} \tag{6}$$

WITH

Y_{ijt}	:	average years of schooling 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[11]
ζ_i	:	age fixed effects
η_j	:	province fixed effects
ε'_{ijt}	:	idiosyncratic error

Column 1 of Table IV gives the OLS results with C_j defined as $WEMI_j$. The estimated coefficient $\overset{\curlywedge}{\beta}_1$ is negative and highly significant, at -0.59. Since the conflict indices are rescaled to the interval $[0,1]$, this value can be interpreted as the change in the number of years of education when moving from zero to maximum conflict intensity.

Insert Table IV about here

The estimate for β_1 may be biased due to reversed causality, i.e. the intensity of violence 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.

I 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 & Stam, 2009)[12]. The exogenous character of the border stems from colonial history, as it was fixed following a compromise agreement by European nations (Department of State, 1965).

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 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, which makes it a relevant instrument for the intensity of ethnic violence.

To instrument for the interaction terms that include $WEMI$, I follow the procedure proposed in Wooldrige (2000, p.236); first constructing predicted values of $WEMI$ by regressing $WEMI$ on the included and the excluded instruments (column 3, Table IV); then using the interaction terms between the predicted $WEMI$, the post-treatment year, and the young age cohort as additional identifying instrument in the first stage of the IVE (columns 4-7). The first stage results demonstrate the relevance of both instruments[13]. Column (2)

[12] Previous work on Rwanda and elsewhere has used similar instruments. Miguel & Roland (2011) use the distance to the meridian that distinguishes North from South Vietnam as an instrument for bombing intensity in Vietnam, while Akresh & 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.

[13] 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 battles intensifying as the RPF

of Table IV reports the second stage IV results, which are very similar to the OLS results. Table AI in Appendix shows that the results remain basically the same when using different specifications of $WEMI$.

Different measures, different results

As evident from Table I above, there exist a number of studies on the micro-economic consequences of violence in Rwanda. For the construction of conflict intensity measures, these studies have borrowed from several sources, including event data, the Genodynamics project (Davenport & Stam, 2009), the 1991 Rwandan population census, Yale genocide studies and the records from the Gacaca (the transitional justice system for genocide suspects). I compiled the previously used measures into one database, and constructed a number of additional measures from the same data sources as well as from ACLED (Raleigh et al., 2010). A detailed description of all 14 alternative measures can be consulted in the Appendix. Table V gives an overview.

The 14 measures include four province-level dummies of conflict intensity (C1-C4 in Panel A of Table V); six continuous province-level measures (C5-C10 in Panel B); and four commune-level measures (C11-C14 in Panel C). These measures are not all positively correlated with WEMI. In particular, the province-level genocide dummy (C1), the province-level proportion of Tutsi (C7), and the province-level share of genocide suspects in the population (C10) correlate negatively with WEMI. This finding may be explained by the genocide-specific character of these measures, omitting other forms of violence, or because these measures are defined at the province level and fail to capture important within province variation in conflict intensity. There is some support for the latter explanation, since, when defined at the commune-level, both the proportion of Tutsi and of genocide suspects correlated positively (instead of negatively) with WEMI.

Insert Table V about here

The last two columns of Table V list the results of the OLS and IV estimates of the coefficient β_1 across the 14 different conflict intensity measures. The estimated DDD coefficients vary considerably across the measures, with significantly negative coefficient estimates in four cases, significantly positive estimates in two cases, and insignificant results in the re-

proceeded.

maining eight cases. Thus, even in a seemingly straightforward application of the impact of armed conflict on schooling, different micro-level conflict measures yield strinkingly different results. Probably, the source of these differences lies in the different degrees of neutrality and spatial detail of these measures. This does not imply that some of these measures are useless, but it does mean that whenever used in an empirical application, one should be clear about what exactly is measured and what is left out, and how this may affect the results.

Conclusion

Besides a steady increase in the number of micro-empirical studies on the legacy of armed conflict, there have been considerable improvements in methodology. In particular, scholars have increasingly devoted attention to identifying rich micro-level measures of conflict intensity (e.g. Restrepo et al., 2006; Raleigh et al., 2010; Brück et al., 2010). This is no coincidence because the identification of micro-level consequences of armed conflict stands or falls with the conflict intensity measure used. So, how should we measure micro-level conflict intensity in order to take a step forward in the study of the legacy of armed conflict? There is no single answer to this question, as one probably needs a combination of various methodologies.

This paper 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 $WEMI$. It is argued that the index is well suited in a context in which different forms of violence took place as well as in the presence of a high direct or indirect death toll. $WEMI$ 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. The findings in the empirical application on schooling in Rwanda suggests that the 1991-2002 schooling deficit in Rwanda can not only be attributed to genocide, but also to other events in the Rwandan conflict cycle, in particular the 1995-1998 counterinsurgency.

Repeating the same empirical exercise with a large number of alternative conflict intensity measures yields very different results, suggesting that empirical applications on the legacy of armed conflict should devote due attention to evaluate the conflict intensity measures used and the underlying methodologies, and - whenever possible - test the robustness of the results against the use of different measures.

$WEMI$ 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 larger set of conflict intensity measures, 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.

$WEMI$ 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, 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.

Appendix: Description of 14 alternative conflict intensity measures

This appendix describes 14 conflict intensity measures for Rwanda (denoted C1-C14). These measures are used to produce the results in Table V. Ten (C1-C10) of the 15 measures are at the province level, the remaining four (C11-C14) are at the commune level. The measures rely on six distinct primary data sources.

Event data

C1: '94 genocide dummy

C2: '90-'94 civil war dummy

C3: '94-'98 (counter)insurgency dummy

Justino and Verwimp (2006) rely on event data - in particular news reports and reports of Human Rights Watch - to evaluate the intensity of genocide, civil war and counterinsurgency 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 1990-1994 civil war intensity (Kibungo, Rural Kigali, Ruhengeri and Byumba - C2) and high levels of 1995-1998 (counter)insurgency (Ruhengeri and Gisenyi - C3).

The Genodynamics project

C4: '94 death toll dummy

C5: 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 (C4). In addition, they make use of the province level number of days with killings in the period April-June 1994 (C5). Both measures are taken from the Genodynamics project, 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).

ACLED

C6: Number of battle events

The Armed Conflict Location and Event Data (ACLED) records 322 battle events in Rwanda for the period 1990-2002[14]. 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. The ACLED events record battles between the RPF and FAR, and do not capture the violence against civilians taking place during the genocide. Hence, the recorded events are concentrated in the northern provinces, where civil war and (counter)insurgency took place.

The 1991 population census

C7 & C11: Province and commune level proportion of Tutsi in the population

From the 1991 population census, one can derive the proportion of Tutsi both at the province level (C7) and at the commune level (C11). 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.

Yale Genocide Studies

C8 & C12 : Province and commune level number of mass graves

C9 & C13: 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). I construct two related measures, which are defined at the province level (C8-C9) as well as the commune level (C12-C13): the number of massgraves and the distance to the nearest mass grave. The latter is calculated by overlaying a geo-referenced administrative map with the location of 71 mass graves in Rwanda.

[14]Note that there is also an ACLED dataset recording events for the period 1997-2010 (Raleigh et al., 2010). Here, I use the older version of ACLED since it spans the entire period of civil war in Rwanda.

The gacaca records

C10 & C14: 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 (C10 & C14).

Acknowledgments

I'm grateful to Tilman Brück, Koen Decancq, Geert Dhaene, 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, Richard & Damien de Walque (2008) Armed conflict and schooling: Evidence from the 1994 Rwandan genocide. HiCN Working Papers 47, Households in Conflict Network.

Akresh, Richard; Philip Verwimp & Tom Bundervoet (2011) Civil war, crop failure, and child stunting in Rwanda. *Economic Development and Cultural Change* 59(4): 777-810.

Alderman, Harold; John Hoddinott & Bill Kinsey (2006) Long term consequences of early childhood malnutrition. *Oxford Economic Papers* 58(3): 450-474.

Alesina, Alberto & Roberto Perotti (1996) Income distribution, political instability, and investment. *European Economic Review* 40(6): 1203-1228.

Amnesty International (1997). Rwanda: civilians trapped in armed conflict. Discussion Paper AI Index AFR 47/43/97, Research Department.

Barro, Robert J. (1991) Economic growth in a cross-section of countries. *Quarterly Journal of Economics* 106(2): 407-444.

Bellows, John & Edward Miguel (2009) War and institutions: New evidence from Sierra Leone. *American Economic Review* 96(2): 394-399.

Blattman, Christopher & Edward Miguel (2010) Civil war. *Journal of Economic Literature* 48(1) :3-57.

Brück, Tilman; Patricia Justino, Philip Verwimp & Alexandra Avdeenko (2010) Identifying conflict and violence in micro-level surveys. Technical report IZA DP No. 5067.

Bundervoet, Tom (2007) Livestock, activity choices and conflict: evidence from Burundi. HiCN Working Paper 24, Households in Conflict Network.

Bundervoet, Tom; Philip Verwimp & Richard Akresh (2009) Health and civil war in rural Burundi. *Journal of Human Resources* 44(2): 536-563.

Chamarbagwala, Rubiana & Hilcías E. Morán (2008) The human capital consequences of civil war: Evidence from Guatemala. HiCN Working Paper 59, Households in Conflict Network.

Davenport, Christian & Allan Stam (2009) Rwandan political violence in space and time. Unpublished manuscript.

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

Deininger, Klaus (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. Technical report Washington, D.C., US.

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

Gleditsch, Nils Petter; Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg & Håvard Strand (2002) Armed conflict 1946-2001: A new dataset. *Journal of Peace Research* 39(5): 651-637.

González, María A. & Rigoberto 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. Technical report, National service of Gacaca Jurisdiction.

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

Hill, Kenneth & James T. Trussell (1977) Further developments in indirect mortality estimation. *Population Studies* 31(2): 75-84.

Human Security Report Project (2010) Human Security Report 2009/2010: The causes of peace and the shrinking costs of war. Pre-publication, Forthcoming in print from Oxford University Press.

Jolliffe, Ian T. (2002) *Principal Component Analysis.* Number XXIX in Springer Series in Statistics, Springer, NY.

Justino, Patricia & Philip Verwimp (2006) Poverty dynamics, violent conflict and convergence in Rwanda. HiCN Working Papers 16, Households in Conflict Network.

Kondylis, Florence (2008) Agricultural outputs and conflict displacement: Evidence from a policy intervention in Rwanda. *Economic Development and Cultural Change* 57(1):31-66.

Lai, Brian & Clayton Thyne (2007) The effect of civil war on education, 1980-97. *Journal of Peace Research* 44(3): 277-292.

Leon, Gianmarco (2010) Civil conflict and human capital accumulation: The long term effects of political violence in Peru. Working Papers id:2505 esocialsciences.com.

Li, Dan (2007) The economic effects of civil war: Evidence from Chinese national railroads, 1906-1923. Unpublished manuscript.

Miguel, Edward & Gerard Roland (2011) The long run impact of bombing Vietnam. *Journal of Development Economics* 96(1): 1-15.

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

Raleigh, Clionadh; Andrew Linke, Håvard Hegre & Joakim Karisen (2010) Introducing ACLED: An armed conflict location and event dataset. *Journal of Peace Research* 47(5): 651-660.

Restrepo, Jorge; Michael Spagat & Juan Vargas (2006) The severity of the Colombian conflict: Cross-country datasets versus new micro-data. *Journal of Peace Research* 43(1): 99-115.

Reyntjens, Filip (2009) The Great African War: Congo and Regional Geopolitics, 1996-2006. New York, N.Y.: Cambridge University Press.

Shemyakina, Olga (2011) The effect of armed conflict on accumulation of schooling: Results from Tajikistan. *Journal of Development Economics* 95(2):186-200.

Timaeus, Ian M. (1986) Estimation of mortality from orphanhood in adulthood. *Demography* 28(2):213–227.

Venieris, Yannis & Dipak Gupta (1986) Income distribution and socio-political instability as determinants of savings: A cross-sectional model. *Journal of Political Economy* 96(4): 873-883.

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

Verpoorten, Marijke (2009) Cattle sales in war and peacetime: A study of household coping in Rwanda, 1991-2001. *Journal of Development Economics* 88(1): 67-86.

Verpoorten, Marijke (2010) Detecting hidden violence: The spatial distribution of excess mortality in Rwanda. Discussion Paper 254 LICOS - Centre for Institutions and Economic Performance, K.U.Leuven.

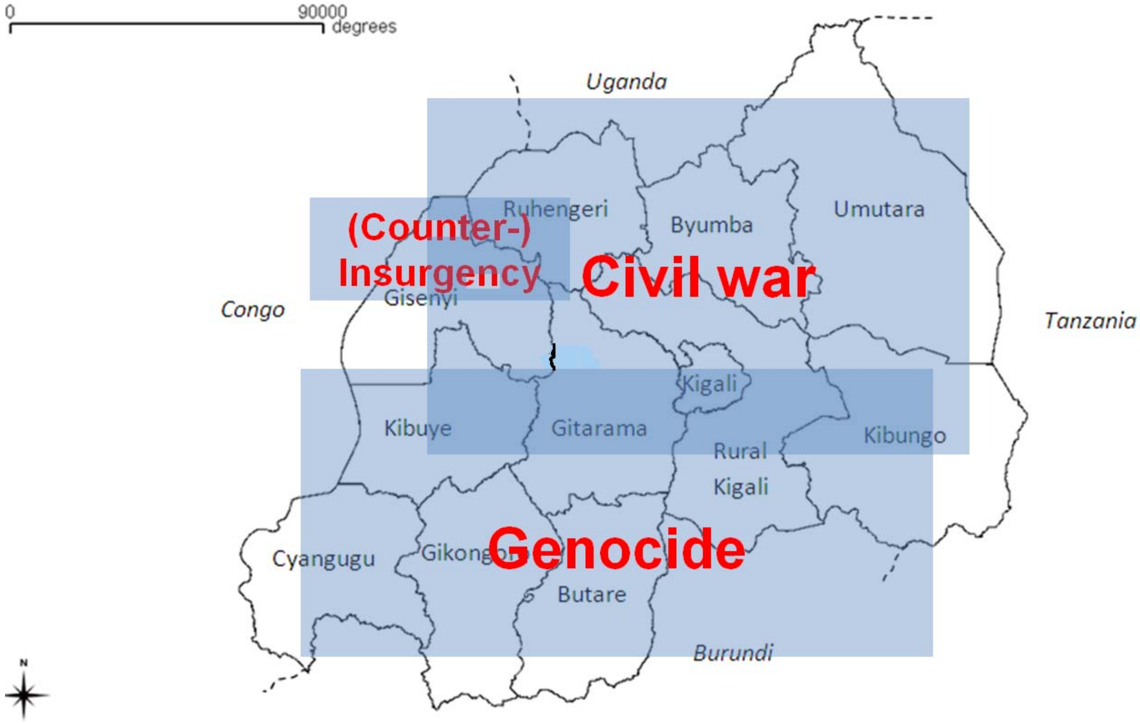
Verpoorten, Marijke & Lode Berlage (2007) Economic mobility in rural Rwanda: A study of the effects of war and genocide at the household Level. *Journal of African Economies* 16(3): 349-392.

Verwimp, Philip & Jan Van 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.

Wooldridge Jeffrey M. (2000). *Econometric Analysis of Cross Section and Panel Data.* MIT Press.

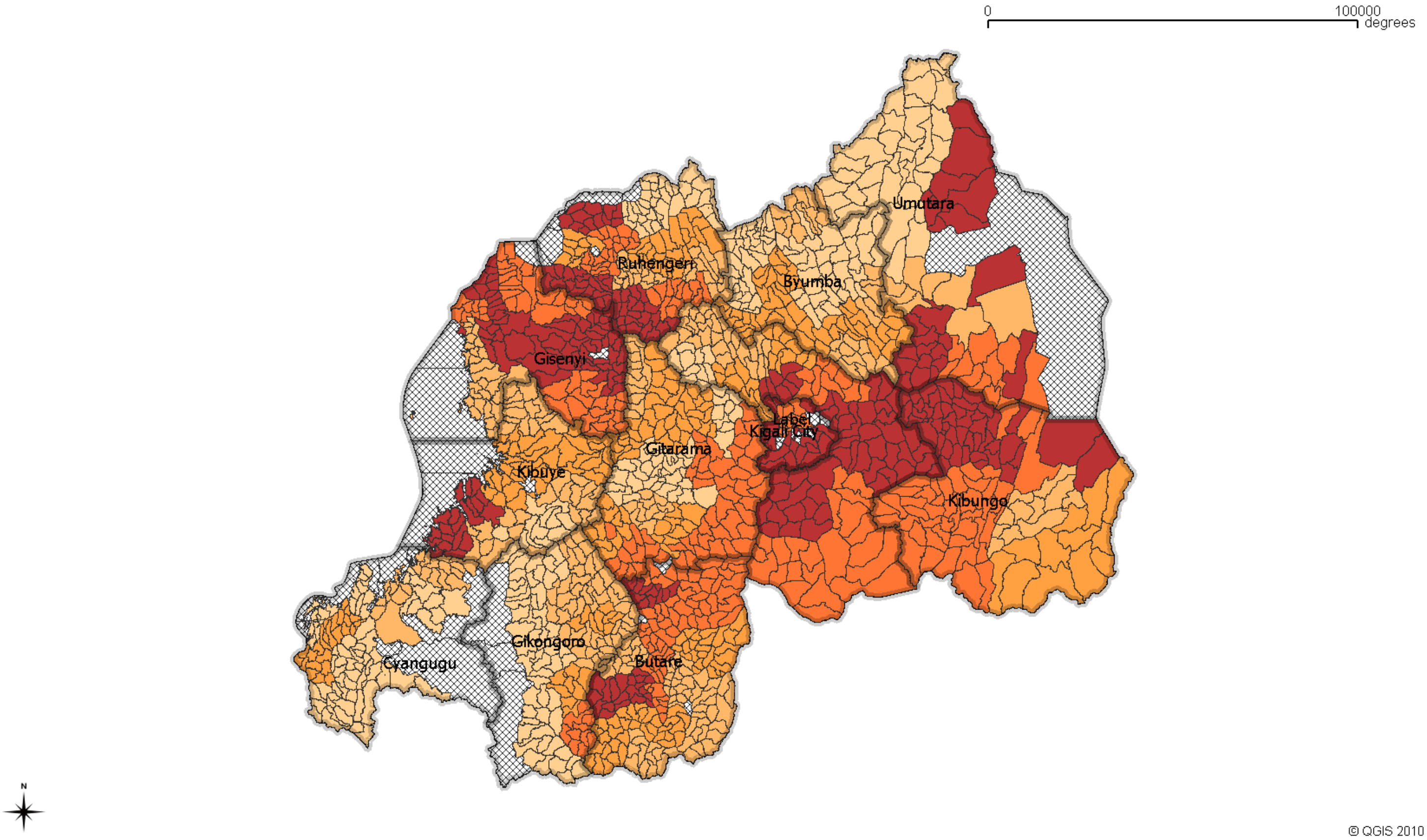
Yanagizawa, David (2010) Propaganda and conflict: theory and evidence from the Rwandan genocide. Unpublished manuscript.

Figure I. Geographical spread of genocide, civil war and (counter-)insurgency



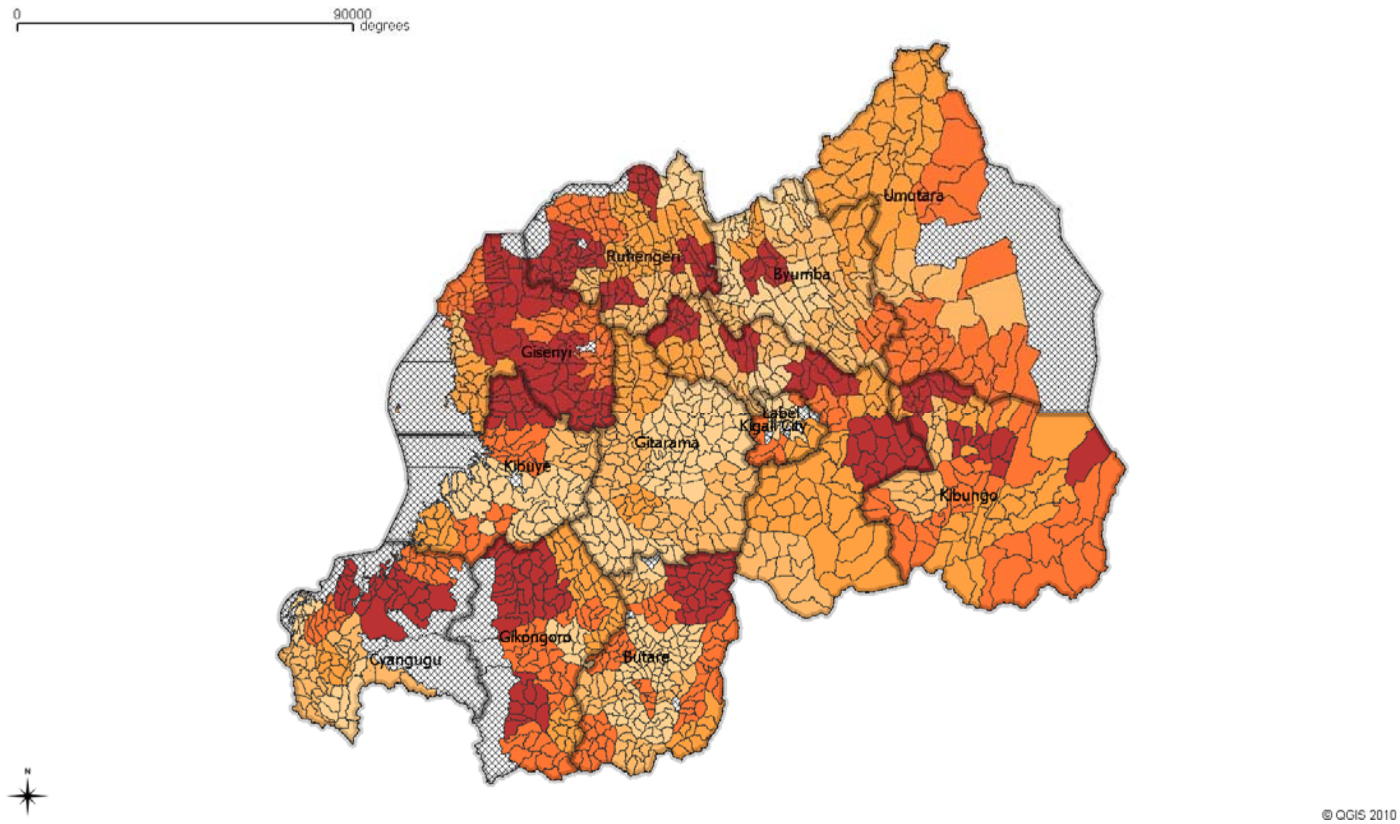
Map taken from shape file of Rwandan provinces; location of different forms of violence based on event data; civil war includes the 1991-1993 and 1994 civil war between the RPF and the FAR; (Counter-)insurgency refers to the 1995-1998 encounters between the RPF and the remains of the FAR and the Interhamwe militia.

Figure II. Map of WEMI quintiles



Top quintile in darkest shade. The map is taken from a shape file of the Rwandan administrative sectors; the quintiles are calculated at the commune level (which is one level above the sector, but for which no shape file exists). The checkered areas are left out of the analysis. They include the national park, forest areas and lakes.

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



Top quintile (= largest schooling deficit) in darkest shade. The map is taken from a shape file of the Rwandan administrative sectors; the quintiles are calculated at the commune level (which is one level above the sector, but for which no shape file exists). The checkered areas are left out of the analysis. They include the national park, forest areas and lakes.

Table I. 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, 2011
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., 2011
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, Nr=20)	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, 2011
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 identification of impact using (I) conflict exposure in time, (II) conflict exposure in space, (III) household conflict experience questions

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

Table II. Commune level 1991-2002 excess mortality proxies for Rwanda (N=145)

Variable	Description	MP (1991 census) mean	st.dev.	MP' (2002 census) mean	st.dev.	WEMP (First difference) mean	st.dev.
Panel A: Mortality proxies used in basline results (excluding migrants)							
Mortality of sons	Boys died/number of boys born	0.199	(0.033)	0.284	(0.037)	0.085	(0.031)
Mortality of daughters	Girls died/number of girls born	0.176	(0.033)	0.244	(0.035)	0.069	(0.030)
Widowhood	Widows (% women >=30)	0.183	(0.030)	0.306	(0.050)	0.123	(0.054)
Double orphanhood	Double orphans (% individuals <30)	0.020	(0.007)	0.051	(0.018)	0.030	(0.017)
Disability	Disabled due to war or genocide (% population)	0.000	(0.000)	0.003	(0.002)	0.003	(0.002)
Panel B: Two of the mortality proxies used in the robustness checks							
Maternal orphanhood	Maternal orphans (% indivdiuals <30)	0.062	(0.015)	0.094	(0.022)	0.032	(0.024)
Paternal orphanhood	Paternal orphans (% individuals <30)	0.122	(0.025)	0.241	(0.050)	0.119	(0.051)

MP and MP' are vectors of pre-war and post-war mortality proxies resp.; WEMP are Wartime Excess Mortality Proxies, calculated as the first difference of MP' and MP. Other mortality proxies that are used in the robustness checks are described in the text but are not summarized in this Table, for reasons of pasimony, and because their summary statistics are very similar to those of the mortality proxies listed in Panel A.

Table III. 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***	-0.995***
	[0.009]	[0.009]	[0.012]

*** significant at 1%. St.errors between brackets. The DD is obtained by estimating Equation (5) for the full sample of indivdiuals. Young cohort = 6-22; old cohort = 23-50; 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 IV. Difference-in-Difference-in-Differences measuring the impact of armed conflict on years of schooling (OLS&IV)

	OLS	IV 2nd stage	IV 1st stage				
Dependent Variable:	Years of Schooling	Years of Schooling	WEMI	WEMI	WEMI*(Young Cohort*CS2002)	WEMI* CS2002	WEMI* Young cohort
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
WEMI * (Young Cohort * CS 2002)	-0.588***	-0.657***					
	(0.117)	(0.125)					
WEMI*CS2002	0.311***	0.384***					
	(0.090)	(0.107)					
WEMI*Young cohort	-0.276	-0.508*					
	(0.224)	(0.290)					
WEMI	0.258*	0.386*					
	(0.154)	(0.214)					
Young Cohort * CS 2002	-0.922***	-0.757***	0.002*	-0.001	0.141***	-0.002	-0.005
	(0.049)	(0.070)	(0.001)	(0.002)	(0.026)	(0.002)	(0.006)
CS 2002	0.528***	0.423***	-0.011***	-0.014***	0.000	0.143***	-0.000
	(0.037)	(0.057)	(0.002)	(0.005)	(0.001)	(0.025)	(0.002)
log(distance to Nyanza)			0.088***	0.088***	0.001	0.001	0.000
			(0.030)	(0.030)	(0.010)	(0.015)	(0.018)
log(distance to Uganda)			-0.065***	-0.065***	-0.001	-0.003	0.001
			(0.025)	(0.025)	(0.007)	(0.012)	(0.014)
predicted WEMI * (Young Cohort * CS 2002)				0.005	0.511***	0.003	0.010
				(0.004)	(0.047)	(0.004)	(0.008)
predicted WEMI*CS2002				0.005	-0.001	0.507***	-0.000
				(0.008)	(0.002)	(0.047)	(0.004)
predicted WEMI*Young cohort				-0.002	0.000	0.000	0.501***
				(0.003)	(0.001)	(0.001)	(0.048)
Individual & Household Level Controls?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Child Age Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Commune level clusters	145	145	145	145	145	145	145
R^2	0.395	0.395	0.463	0.833	0.888	0.856	0.463
F-stat of excl. instrument			9.7	9.53	58.94	53.86	58.81
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Sargan overidentification test		0.199					
		(0.655)					
Cragg-Donald statistic		25747					
Observations	1,074,561	1,074,562	1,074,563	1,074,567	1,074,565	1,074,566	1,074,564

*** p<0.01, ** p<0.05, * p<0.1. The unit of observation is an individual. The robust standard errors are adjusted for clustering within communes and are reported 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. The interaction terms are instrumented as suggested by Wooldrige (2000, p.236); first constructing predicted values of WEMI by regressing WEMI on the included and the excluded instruments (column 3); then using the interaction terms between the predicted WEMI, the post-treatment year and the young age cohort as additional identifying instrument in the first stage of the IVE (columns 4-7).

Table V. Overview of conflict intensity measures for Rwanda and the corresponding DDD estimates

		Data source	Used in	Mean	St. dev.	Correlation with WEMI[2]	Estimated DDD Impact[3]	
							OLS	IV 2nd stage
Panel A: province-level dummies								
C1	Genocide dummy	Even data	Justino & Verwimp, 2006	0.35	0.48	-0.37***	0.168***	0.180***
C2	'90-'94 civil war dummy	Even data	Justino & Verwimp, 2006	0.44	0.50	0.21***	-0.135**	-0.134**
C3	'94-'98 (Counter-)insurgency dummy	Even data	Justino & Verwimp, 2006	0.20	0.40	0.26***	-0.119**	-0.121**
C4	'94 death toll dummy	Genodynamics project	Akresh & de Walque, 2008	0.32	0.46	0.28***	-0.139**	-0.082
Panel B: province-level continuous measures								
C5	Days with killings in April-July '94	Genodynamics project	Akresh & de Walque, 2008	24.22	20.39	-0.01	0.155*	0.163*
C6	Number of battle events	ACLED	N.A.	30.87	18.11	0.07	0.014	0.106
C7	Proportion of Tutsi	1991 popuation census	Akresh & de Walque, 2008	0.08	0.06	-0.10***	0.029	-0.039
C8	Number of mass graves	Yale genocide studies	Akresh & de Walque, 2008	6.28	4.88	0.05***	0.006	-0.016
C9	Distance to mass grave	Yale genocide studies	N.A.	12.29	7.81	-0.13***	0.037	0.092
C10	Genocide suspects(%)	Gacaca records	N.A.	0.07	0.04	-0.06***	0.077	0.176*
Panel C: commune-level measures								
C11	Proportion of Tutsi	1991 popuation census	N.A.	0.12	0.14	0.18***	-0.162	-0.189
C12	Mass grave	Yale genocide studies	N.A.	0.47	0.81	0.20***	-0.354**	-0.298**
C13	Distance to mass grave	Yale genocide studies	N.A.	12.29	10.17	-0.22***	0.097	0.091
C14	Genocide suspects (%)	Gacaca records	Yanagizawa, mimeo[1]	0.07	0.05	0.10***	0.018	0.216*

[1] Yanagizawa uses the share of genocide suspects in the population as a dependent variable rather than as an explanatory variable; [2]the correlation coefficients are calculated from the individual-level dataset, but they are similar when calculated at the commune level (although significance decreases); [3] The OLS and IV DDD effects are calculated using the same specification as in Table 4, replacing WEMI with the different alternative measures.

Table AI. Robustness checks for WEMI and the DDD estimates

Wartime Excess Mortality Index	Correlation coefficient with WEMI	Estimated DDD Impact OLS	IV 2nd stage
Basline WEMI	1	-0.588***	-0.657***
		(0.117)	(0.125)
Adding distance to mass grave	0.997	-0.596***	-0.664***
		(0.119)	(0.131)
Including migrants	0.998	-0.563***	-0.643***
		(0.115)	(0.121)
Age limit: 25 instead of 30	0.998	-0.576***	-0.647***
		(0.116)	(0.123)
Age limit: 35 instead of 30	0.998	-0.604***	-0.650***
		(0.118)	(0.124)
Maternal orphanhood instead of double orphanhood	0.993	-0.582***	-0.626***
		(0.115)	(0.117)
Paternal orphanhood instead of double orphanhood	0.991	-0.522***	-0.568***
		(0.102)	(0.111)
Drop mortality sons	0.987	-0.602***	-0.708***
		(0.123)	(0.138)
Drop mortality daughters	0.985	-0.587***	-0.703***
		(0.113)	(0.138)
Drop widowhood	0.986	-0.595***	-0.661***
		(0.126)	(0.129)
Drop double orphanhood	0.988	-0.495***	-0.584***
		(0.110)	(0.115)
Drop disability due to war/genocide	0.986	-0.588***	-0.631***
		(0.111)	(0.118)

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. The unit of observation is an individual. The robust standard errors are adjusted for clustering within communes and are reported between brackets. The regressors are the same as those included in the regression model underlying the results in Table 4. The estimated DDD effect corresponds with the estimated coefficient on the interaction term WEMI * (Young Cohort * CS 2002).

Table AII. District level 1991-2002 excess mortality proxies for Tanzania (N=113)

Variable	Description	MP (1988 census) mean st.dev.	MP' (2002 census) mean st.dev.	WEMP (First difference) mean st.dev.
Mortality of sons	Boys died/number of boys born	0.206 (0.041)	0.175 (0.041)	-0.031 (0.019)
Mortality of daughters	Girls died/number of girls born	0.189 (0.039)	0.160 (0.039)	-0.030 (0.019)
Widowhood	Widows (% women >=30)	0.166 (0.038)	0.170 (0.048)	0.004 (0.034)
Maternal orphanhood	Maternal orphans (% individuals <30)	0.053 (0.010)	0.068 (0.020)	0.015 (0.019)

MP and MP' are vectors of mortality proxies for 1988 and 2002 resp.; WEMP are 1988-2002 Excess Mortality Proxies, calculated as the first difference of MP' and MP.

ONLINE APPENDIX TO THE ARTICLE

"Measure for Measure:

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

Marijke Verpoorten*

September, 2011

Abstract

WEMI, the measure developed in the main article, can be the basis for the construction of a multi-dimensional conflict intensity measure. For example, when adding data on a specific form of violence to the set of "neutral" excess mortality proxies, a two-dimensional measure can be obatined. For the case of Rwanda, I develop a two-dimensional index by augmenting the 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-related 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).

Extension: distinguishing between different forms of violence

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 excess mortality proxies subjected to PCA with rich information on genocide intensity. This information is taken

*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

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 (2010) 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, projected backwards from 2005 to 1994[1], 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 (% commune-level 1994 population)
- (GP_{j2}) Genocide suspects accused of killings or other serious physical assaults (% commune-level 1994 population)
- (GP_{j3}) Genocide suspects accused of looting or other offences against property (% commune-level 1994 population)
- (GP_{j4}) Widowed genocide survivors (% commune-level 1994 population)
- (GP_{j5}) Orphaned genocide survivors (% commune-level 1994 population)
- (GP_{j6}) Disabled genocide survivors (% commune-level 1994 population)

Table 1 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

[1] Since most of the gacaca data were collected in 2005, I project the numbers backwards to 1994. For example $GP_{j1} : \frac{(category\_1\_suspects_{i2005})(1-d_n)^{-11}}{population_{i1994}}$ with d_n the annual death rate among the adult population estimated from the 2000 Rwandan DHS survey (Timaeus & Jasseh, 2004)

linear combinations:

$$
\begin{aligned}
PC1_j &= [0.23, 0.19, 0.12, 0.16, 0.23]' \times e_j + [0.39, 0.39, 0.31, 0.39, 0.39, 0.33]' \times g_j \qquad (1) \\
PC2_j &= [0.40, 0.38, 0.46, 0.46, 0.31]' \times e_j + [-0.15, -0.19, -0.25, -0.18, -0.15, -0.08]' \times g_j \\
PC3_j &= [0.20, 0.33, 0.04, 0.02, -0.49]' \times e_j + [0.19, 0.27, 0.45, -0.24, -0.20, -0.45]' \times g_j \\
PC4_j &= [-0.33, -0.57, 0.39, 0.40, 0.15]' \times e_j + [0.25, 0.02, 0.32, -0.19, -0.19, -0.06]' \times g_j
\end{aligned}
$$

Combined, these four PCs explain 86% of the total variance in the underlying set of 11 variables. However, $PC3_j$ and $PC4_j$ 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 ($CEMI$)[2].

Figures 1a and 1b 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 I in the main article. $CEMI$ 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 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 & Stam, 2009).

I perform a series of robustness checks for GEMI and CEMI. First, I vary the excess mortality proxies as I did for WEMI in the main article. Second, I exclude the genocide

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

proxies case by case. The correlation coefficients between the alternative and the baseline indices are all above 0.95, as is shown in Table A1.

Difference-in-Difference-in-Difference estimates: GEMI, CEMI and the 1991-2002 schooling deficit

The following equation is estimated:

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

WITH

Y_{ijt}	:	average years of schooling 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[3]
ζ_i	:	age fixed effects
η_j	:	province fixed effects
ε_{ijt}	:	idiosyncratic error,

with C_j defined as $GEMI_j$ and $CEMI_j$. Table 2 gives the results. The estimated coefficient $\hat{\beta}_1$ is negative and highly significant, with respective values of -0.13 and -0.63. These results suggest that, compared to $GEMI_j$, the impact of $CEMI_j$ on schooling is more severe. In column (3), $GEMI_j$ and $CEMI_j$ are jointly included, which yields very similar point estimates to those of columns (1) and (2) because $GEMI_j$ and $CEMI_j$ are orthogonal vectors (which is a feature of PCA).

Compared to the OLS results, the IV results for β_1 in columns (4) and (5) are very similar. However the IV estimate 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. Note that this is not an issue in the OLS results 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 (6), the coefficient on $\widehat{GEMI}_j$ turns again negative and significant, but still remains below below the estimated impact of $\widehat{CEMI_j}$.

Robustness checks

Both the OLS and IV estimate for β_1 remain qualitatively the same after correcting for survival bias, i.e. when adding the distance to the nearest mass grave to the set of variables subjected to PCA.Looking at the quantitative effect, the estimated OLS (IV) coefficient on $GEMI$ increases by 13.8%, whereas the difference is negligibly small for $CEMI_j$. 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 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). Finally, I replicate the results using different definitions for the young and old age cohort. All robustness checks yield results that are qualitatively the same as those presented in Table 2, and they can be obtained on request.

References

Akresh, Richard & Damien de Walque (2008) Armed conflict and schooling: Evidence from the 1994 Rwandan genocide. HiCN Working Papers 47, Households in Conflict Network.

Davenport, Christian & Allan Stam (2009) Rwandan political violence in space and time. Unpublished manuscript.

Government of Rwanda (2005) Report on data collection: annexes. Technical report, National service of Gacaca Jurisdiction.

Jolliffe, Ian T. (2002) *Principal Component Analysis.* Number XXIX in Springer Series in Statistics, Springer, NY.

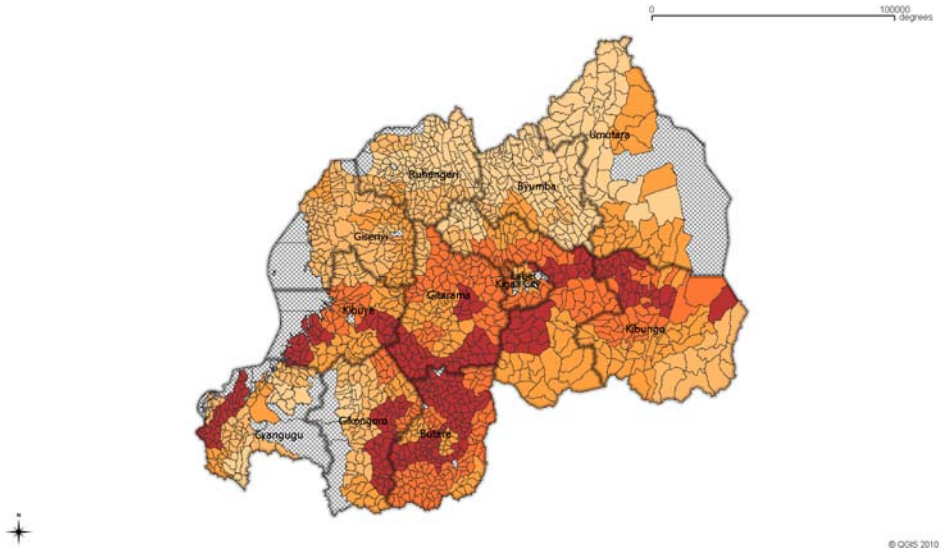
Justino, Patricia & Philip Verwimp (2006) Poverty dynamics, violent conflict and convergence in Rwanda. HiCN Working Papers 16, Households in Conflict Network.

Verpoorten, Marijke (2010) 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.

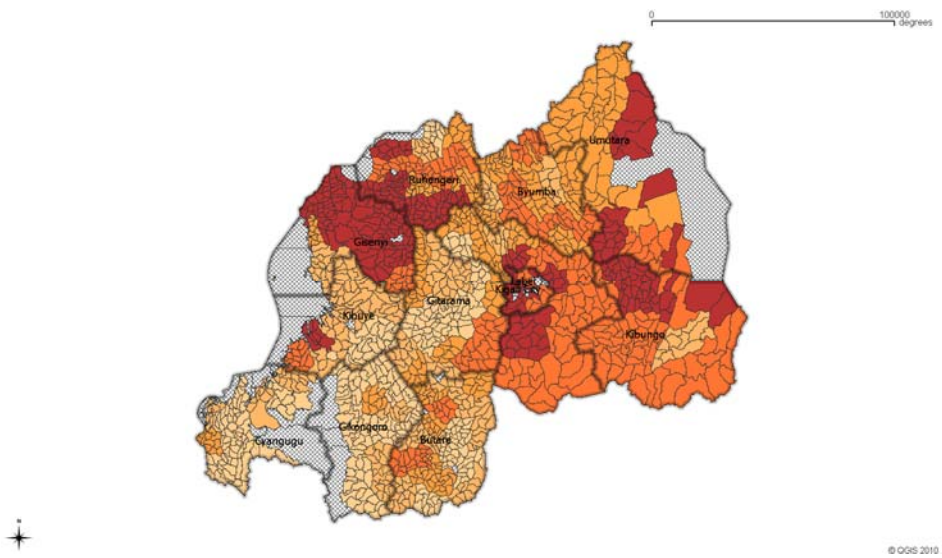
Timaeus, Ian & Momodou Jasseh (2004) Adult Mortality in Sub-Saharan Africa: Evidence from Demographic and Health Surveys. *Demography* 41:757–72.

Figure 1: Map of GEMI and CEMI quintiles

1a. GEMI



1b. CEMI



Top quintile in darkest shade. The map is taken from a shape file of the Rwandan administrative sectors; the quintiles are calculated at the commune level (which is one level above the sector, but for which no shape file exists). The checkered areas are left out of the analysis. They include the national park, forest areas and lakes.

Table 1. 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%

The proxies are taken from the 2005 records of the transitional justice courts (gacaca), and projected backward to 1994; the 1994 population is projected forward from the 1991 population census using commune level 1978-1991 population growth rates.

Table 2: Difference-in-Difference-in-Differences measuring the impact of armed conflict on years of schooling

Dependent Variable: Years of Schooling	OLS			IV 2nd stage		
	(1)	(2)	(3)	(4)	(5)	(6)
GEMI * (Young Cohort * CS2002)	**-0.131*** **		**-0.113** **	0.012		**-0.145*** **
	[0.051]		**[0.051]**	[0.053]		**[0.055]**
GEMI*CS 2002	0.073*		0.067	0.008		0.096**
	[0.042]		[0.042]	[0.044]		[0.046]
GEMI*Young cohort	-0.360***		-0.361***	-0.727***		-0.872***
	[0.036]		[0.036]	[0.000]		[0.040]
GEMI	0.438***		0.492***	1.700***		2.222***
	[0.031]		[0.031]	[0.041]		[0.046]
CEMI * (Young Cohort * CS2002)		**-0.629*** **	**-0.609*** **		**-0.486*** **	**-0.506*** **
		[0.055]	**[0.055]**		**[0.041]**	**[0.042]**
CEMI*CS 2002		0.346***	0.319***		0.301***	0.299***
		[0.046]	[0.045]		[0.034]	[0.035]
CEMI*Young cohort		-0.150***	-0.160***		-0.267***	-0.437***
		[0.039]	[0.039]		[0.030]	[0.030]
CEMI		0.069**	-0.039		0.030	-0.713***
		[0.034]	[0.035]		[0.040]	[0.044]
Young Cohort * CS 2002	-1.128***	-0.922***	-0.899***	-1.142***	-0.924***	-0.808***
	[0.021]	[0.023]	[0.029]	[0.029]	[0.020]	[0.038]
CS 2002	0.637***	0.521***	0.515***	0.551***	0.514***	0.341***
	[0.018]	[0.019]	[0.024]	[0.024]	[0.017]	[0.031]
Individual & Household Level Controls?	Yes	Yes	Yes	Yes	Yes	Yes
Child Age Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes
Province Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,074,561	1,074,561	1,074,561	1,074,561	1,074,561	1,074,561

*** p<0.01, ** p<0.05, * p<0.1. The unit of observation is an individual. The robust standard errors are adjusted for clustering within communes and are reported 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 3: First stage of IVE

Dependent Variable:	GEMI	CEMI
Excluded instruments:		
log(distance to Nyanza)	-0.082**	-0.018
	[0.037]	[0.034]
log(distance to Uganda)	0.052*	0.067**
	[0.028]	[0.026]
Included instruments:		
Household Level Controls	Yes	Yes
Province Fixed Effects?	Yes	Yes
Observations	1,074,561	1,074,561

*** p<0.01, ** p<0.05, * p<0.1; The unit of observation is an individual. The robust standard errors are adjusted for clustering within communes and are reported between brackets. The control variables are the same as in Table 2.

Table A1. Robustness checks for GEMI and CEMI

	Correlation coefficient with GEMI	Correlation coefficient with CEMI
Panel A: Alternative excess mortality proxies		
Adding distance to mass grave	0.994	0.999
Including migrants	0.999	0.998
Age limit: 25 instead of 30	0.999	0.998
Age limit: 35 instead of 30	0.998	0.996
Maternal and paternal orphans	0.990	0.959
Drop WEMP 1	0.980	0.970
Drop WEMP 2	0.988	0.975
Drop WEMP 3	0.995	0.981
Drop WEMP 4	0.991	0.979
Drop WEMP 5	0.985	0.978
Panel B: Alternative genocide proxies		
Drop GP 1	0.966	0.971
Drop GP 2	0.956	0.960
Drop GP 3	0.973	0.972
Drop GP 4	0.958	0.965
Drop GP 5	0.966	0.972
Drop GP 6	0.987	0.995