

Understanding Civil Wars: We Need Better Data

by Juan F. Vargas

[Back to Previous Page](#)

Introduction

Recently, the Environmental Assessment Institute of Denmark organized a forum of experts including some Nobel laureates in Economics. The aim of the so called Copenhagen Consensus was to commission leading scholars to assess the greatest challenges faced by humanity. The panel of experts examined the estimates and prioritized the challenges offering the best potential cost-benefit opportunities.¹

One of the challenges considered in the Consensus was the increasing incidence of civil wars. Indeed, there is now a fair amount of literature on how civil wars hinder economic and human development. This is also the view of multilateral institutions such as the World Bank, which gives growing importance to the analysis of causes and consequences of civil wars in developing countries.²

The experts of the Consensus unanimously agreed with the importance of civil wars as a major threat to development. Nevertheless, they actually omitted civil wars from their list of priorities, pleading insufficient information. What sort of information is there available on the nature and dynamics of civil wars and why is it insufficient?

The existing data

One of the first and most influential quantitative works on conflicts is the Correlates of War project (hereafter COW) described in the pioneering book by David Singer and Melvin Small.³ COW was the first long-term cross-country dataset on the incidence and intensity of both inter- and intra-state conflicts. Constantly updated, its use in quantitative analyses of war is still widespread, although nowadays other datasets are available. Alternatives are the Civil War Termination data, the datasets built respectively by Michael Doyle and Nicholas Sambanis and by James Fearon and David Laitin, and the Uppsala/PRIO dataset from the Department of Peace and Conflict Research at Uppsala University and the International Peace Research Institute at Oslo. But to varying extents all of the later datasets build on COW.

The development of these cross-country datasets has supported a recent boom in the empirics of conflict, expanding our understanding of civil war and supporting policy advice on how to prevent and overcome these conflicts. However, a disturbing question is whether this advice rests on a weak empirical base, as implied by the conclusion of the Copenhagen Consensus. The available econometric findings have generated stimulating but inconclusive debates, and it may well be that the quality of the data is at fault. Surprisingly enough, until very recently no one seems to have posed the question: how good is the data we rely on?

Quality of standard cross-country datasets

Despite their common origins, there is considerable variety among cross-country datasets, beginning with the range of definitions of the object of study. Most of them include a measure of intensity (number of battle-related killings) but some omit intensity measures entirely and limit themselves to listing different conflicts and the time span in which they took place. Among those that do include a conflict intensity number, very few provide time series and when they do they give quite wide ranges. Most datasets just give aggregate numbers without underlying time series. In all cases, the numbers are often poorly documented, which makes it difficult to place great confidence in them.

In a recent paper, coauthored with Jorge Restrepo and Michael Spagat of the Royal Holloway College of the University of London,⁴ we discuss these issues and test the quality of the cross-country datasets. Here I

summarize our simple approach.

Over the past two years, we have developed a general methodology for the in-depth measurement of conflict activity in a single conflict. We have applied this methodology to the Colombian civil war and the result of this effort is a detailed, high frequency time-series dataset (hereafter RSV) that covers more than 21,000 conflict related events over the period 1988-2003. For every event we record the date and the place of occurrence (at the level of the township); whether there was a clash between two or more forces or a one sided, uncontested attack (in which case we distinguish the type of attack and the group responsible for it); and the number of killings and injuries.⁵ The data provides a detailed long-term picture of the temporal and spatial dynamics of the conflict as well as the evolution of the various conflict activities and their impact in terms of casualties. In building the dataset, we have greatly benefited from the efforts of the Colombian NGOs Centro de Investigación y Educación Popular and the Comisión Intercongregacional de Justicia y Paz, who publish *Noche y Niebla*, a quarterly periodical that lists events of political violence gathered from a large network of priests and collaborators as well as from over 25 national and regional newspapers.⁶ We complement this source with press reports and code it into a dataset after applying our methodological filter: We focus merely on civil war dynamics rather than the broader concept of “political violence” not necessarily connected with the conflict.

We evaluate the quality of the cross-country datasets on civil war by comparing their Colombia figures with those of RSV. The latter can be considered a “control dataset” for a sample of the former. Obviously, it is a sample of one, but in the short run it is the only feasible sample, given the high cost of building datasets with the level of detail and the degree of care that RSV applies to Colombia. At this stage we do not know the extent to which our conclusions can be generalized, i.e., whether we have sampled an outlier.

Our comparison suggests that the cross-country datasets have significant quality problems. There are two main issues to point out. First, there is a tendency to underestimate the intensity of the Colombian civil war in terms of the casualties it produces. This finding is especially meaningful because it is difficult to argue that RSV, which rules out all non-conflict-specific violent manifestations (e.g. criminality figures), overestimates the yearly number of killings.

We compare the annual averages for killing rates in Colombia of RSV and 12 of the most important cross-country datasets of civil war. These averages are significantly below the RSV figure for all but four datasets. Of these four, two overestimates are actually very close to the RSV figures. The remaining two datasets provide ranks in which RSV lies, but the intervals are particularly wide.

The exercise with the annual averages is necessary since very few datasets provide actual yearly data. However, when possible we also compare RSV with the datasets that have time series. Almost all of these report wide ranges, making this comparison sometimes ambiguous. In spite of this, the majority of the estimates are unambiguously large underestimates of the annual intensity of the Colombian civil war compared to RSV. The unique case in which there is an overestimation in one year appears to be an error in the respective cross-country dataset. In the cases when the figures are compatible with RSV, the ranges suggested by the datasets are very wide. For 2002, for instance, one dataset underestimates the death rate by 500 in one case and over 2700 in another.

The second main finding is that no dataset captures the actual dynamics of the civil war. In fact, the most prominent trend in the recent evolution of the Colombian war, namely its significant upsurge after 1996, is unreported in the cross-country information.

The future of civil war research

Understanding the nature, the dynamics and the consequences of increasingly common civil wars is necessary to provide urgent public policy advice. Over the last 20 years, the effort of quantitative researchers and social scientists has focused on the development and use of cross-country datasets. This approach has been useful and insightful. But civil wars are still a black box. Significant further progress in civil war research will require improvements and extensions of existing datasets and the development of new ones along the lines of the micro-dataset approach described in this article.

NOTES

1. For more, see www.copenhagenconsensus.com.
2. See the 2003 World Bank Policy Report. Breaking the Conflict Trap: Civil War and Development Policy. Online at <http://econ.worldbank.org/prr/CivilWarPRR/text-26671/>
3. Resort to Arms: International and Civil Wars, 1816-1980. Beverly Hills, CA: SAGE, 1982.
4. The Severity of the Colombian Conflict: Cross-Country Datasets vs. New Micro Data. Online at http://personal.rhul.ac.uk/pkte/126/Pages/research_on_colombia.htm
5. See a detailed description of the dataset in Restrepo, Spagat and Vargas' The Dynamics of the Colombian Civil Conflict: A New Dataset. Online at http://personal.rhul.ac.uk/pkte/126/Pages/research_on_colombia.htm
6. The latest issues of the periodical are online in Spanish at www.nocheyniebla.org.

[Back to top](#)

Juan F. Vargas is completing his PhD in Economics at Royal Holloway University of London.

Economists for Peace and Security
<http://www.epsusa.org>