

Measurement validity in historical conflict data: Comparing datasets from Spain

Francisco Villamil*

September 19, 2021

Abstract

Conflict research usually suffers from data availability problems, which sometimes motivates the use of proxy variables for violent events. But since they are usually the only alternative to measure violence patterns, there is not ground-truth data to compare them to. This limitation explains why there are no studies assessing their validity. This research note exploits a case where there are two sources on political violence: the Spanish Civil War. Comparing georeferenced mass graves and direct records of victimization, I show that the differences between these two datasets are not random but respond to different data generation processes, introducing important biases. Results highlight the need for a more careful assessment when using proxy variables for political violence.

*Carlos III–Juan March Institute, Universidad Carlos III de Madrid. francisco.villamil@uc3m.es

1 Introduction

Following a general shift in the political science literature, during the last few years conflict research has paid increasing attention to the problem of causal identification. However, this methodological shift has not been matched with a similar concern about measurement. Data on patterns of political violence are often hard to collect or find. Perhaps because of that, measurement problems have not been discussed as thoroughly as causal identification strategies.

The literature has partly moved away from macro-level empirical analyses at the global level to an increase of single-case analyses where it is possible to test a causal mechanisms. Two trends are related to this move. First, the emergence of a research agenda on the micro-dynamics of conflict ([Kalyvas, 2008](#)) spearheaded a shift of attention to micro-level variation found within a single country or region. Second, a renewed interest in historical conflicts motivated the search for cases where it is possible to quantitatively identify a causal effect, particularly when studying the long-term consequences of political violence ([Walden and Zhukov, 2020](#)). All these trends explain the use of alternative data sources on political violence, such as archives or proxy variables.

Obtaining reliable data on patterns of political violence is difficult. This limitation is particularly common to almost all studies that focus on violence against civilians ([Balcells and Stanton, 2021](#)), because these events usually take place in low-information environments where actors have incentives to cover up victims, leading to multiple bias in the count of fatalities ([Krüger et al., 2013](#)). Even in relatively modern conflicts, where victims or relatives are still alive, there are important data limitations ([Armstrong et al., 2020](#)). Innovative forms of data collection do not offer a perfect solution either. Event datasets based on media reports have been criticized ([Baum and Zhukov, 2015; Weidmann, 2016](#)), especially when they are machine-coded ([Hammond and Weidmann, 2014](#)), and geographic coding decisions in manually collected data are prone to be biased ([Doughlass and Harkness, 2018](#)). The scarcity of data sometimes motivates the use of proxy measures. For instance, [Kasara \(2016\)](#) uses IDPs and satellite images of fires as proxies

for violence in Kenya. Especially when looking for spatially disaggregated data on past conflicts, these proxies might be the only option available.

Previous works have discussed the use of both archival data and identification strategies on historical conflicts. [Kocher and Monteiro \(2016\)](#) criticize natural experiments in quantitative historical research that do not rest on a thorough qualitative validation. [Balcells and Sullivan \(2018\)](#) discuss the findings of studies that rely on conflict archival data and propose a set of practical recommendations when using this type of data sources. Yet, no previous work discusses the problem of measurement validity in conflict research, particularly in studies that focus on historical conflicts. Partly, this gap exists because there is often no ground-truth data available to compare violence datasets with.

In this research note, I try to fill this gap. I focus on case of the Spanish Civil War, where two alternative data sources on political violence are available and both have been used in previous research ([Balcells, 2017](#); [Barceló, 2018](#); [Tur-Prats and Valencia Caicedo, 2020](#)). In particular, I compare data on the location of mass graves with regional datasets on local patterns of political violence collected by historians from archives and civil registries. Using both simple correlations and spatial regression models, I show that the relationship between mass graves and wartime victimization at the local level varies greatly depending on the period and region. This results suggests not only that mass graves are not a reliable measure of wartime violence, but also show that the use of such a proxy variable suffers from systematic biases. The empirical evidence comes from a single case, but its lesson applies to any empirical study using historical data sources.

This research note contributes to the emerging body of research on historical conflict processes, motivating a more careful discussion of the data sources and the potential biases in the way they were created. Rather than criticizing previous works, its aim is to offer a cautionary tale about the use of proxy data in conflict research. Particularly when focusing on historical phenomena, the results strongly recommend the need for a careful consideration of indirect measures of political violence.

2 Measuring wartime violence in Spain

There are two sources of local-level data on patterns of violence during the Spanish Civil War. First, in some regions of Spain, different research projects have compiled detailed lists of wartime victims either at the individual level, including information on where they lived, died, or were born, or at the level of municipalities, indicating the number of killings in each location. There are varying degrees of detail or coverage depending on the project. While some projects cover all victims by both fighting sides, others are limited to civilian victimization by Francoist forces. For instance, the project for Asturias is arguably the most detailed one: it covers every type of killing by both Francoist and Republican forces and includes details on the date, location, and characteristics of the victim (García et al., 2011). All these projects, however, have a relatively good quality since these lists were created by triangulating multiple primary sources, such as death records in the local civil registries with historical documents and testimonies. Here I include data from seven Spanish regions: Galicia, Asturias, the Basque Country, Aragón, Catalunya, Albacete (Castilla-La Mancha), and Badajoz (Extremadura).¹ Although there are available lists of victims for other regions, many of them do not have the same degree of quality since they usually consist of a list of those deaths or disappearances claimed by the victims' relatives or friends.

Second, as part of the Historical Memory Law, the Spanish Government has published an online database with all the mass graves located throughout the country. It includes information on the location of each mass grave and the number of identified and not identified victims.² An important advantage of this database is that it covers the whole country. Thus, the location of mass graves could be used as a proxy for local wartime victimization, getting data that covers the entire country. The problem with this strategy, as with the use of proxies for political violence in general, is that the relationship between both phenomena can be systematically biased.

¹I include in the Appendix more details about each dataset on wartime killings.

²Data was automatically retrieved using an R script from the online database website, which can be found at <https://www.mpr.gob.es/memoriademocratica/mapa-de-fosas/Paginas/buscadormapafosas.aspx> (accessed 17/03/2021).

In this note I compare these two data sources, shown in figure 1. In particular, I compare the local-level number of Francoist violence against civilians, linking them to their municipality of residence, with the number of mass graves and the total number of victims in mass graves in each municipality. In order to explore biases explained by the different data generating process, I also compare these datasets restricting the sample to each region, looking only at the number of Francoist killings in different years, or distinguishing between irregular killings and summary executions. Where available, I also use the total number of victims from both sides.

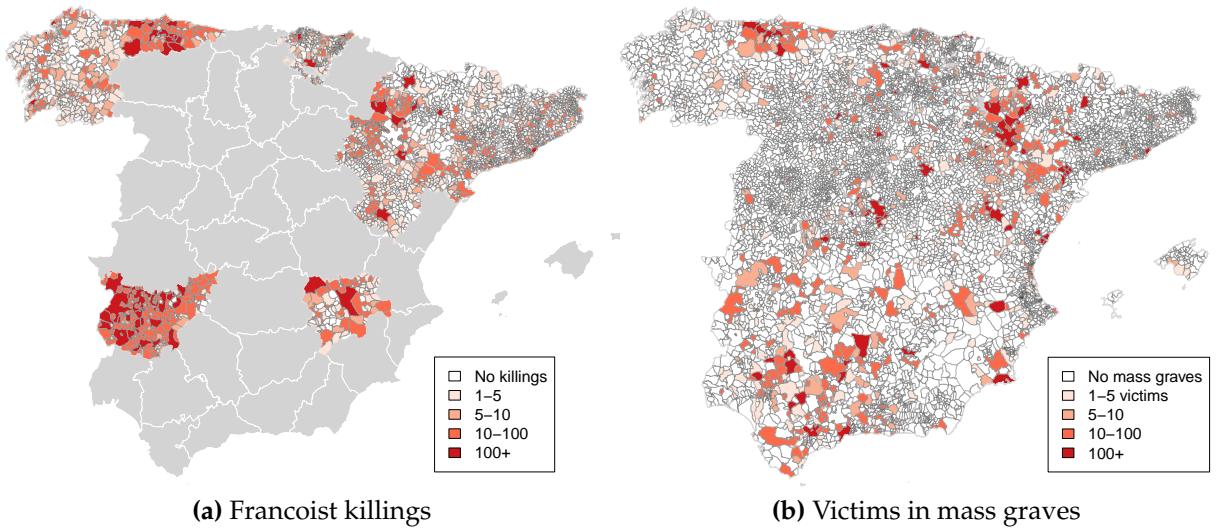


Figure 1: Local-level Francoist killings and mass grave data

Moreover, I also make a detailed comparison for the data in Asturias. In this region, the two datasets have the most quality, given that the same team of researchers compiled comprehensive lists of both mass graves and wartime killings. Limiting the sample to this region, even if the sample is greatly reduced, allows me to assess the correlation between mass grave data and different types of wartime political violence. Keeping measurement error at minimum, these correlations show a fundamental reason why proxies for violence need to be used cautiously: the data generating process for the two of them is different.

Finally, beyond simple correlations, I use spatial regression models to account for spatial correlation and non-local predictor effects. In particular, the location of mass graves could be linked to wartime killings in a contiguous municipality.

3 Comparing datasets

Figure 2 shows the correlation between the number of Francoist killings between 1936 and 1942 and both the number of mass graves and the number of victims in mass graves in each municipality, limiting the sample to those regions where both datasets are available. All the variables are shown in logarithmic form to minimize the influence of outliers.³

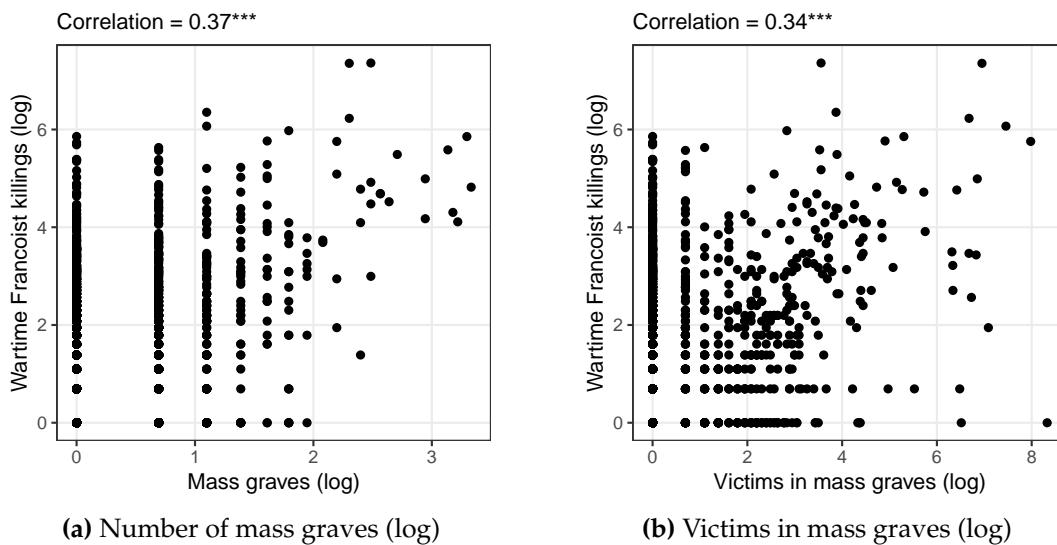


Figure 2: Local-level Francoist civilian killings (log) and mass grave data

The overall correlation of local victimization with the number of mass graves is relatively small (0.37). This number becomes slightly smaller when comparing with the total number of victims found in mass graves (0.34), which is probably explained by the fact that many mass graves (around 50%) have not yet been exhumed, and many of these do not have a definite number of victims.

Table 1 compares the two datasets by region. The correlation coefficients vary strongly. In Asturias the two datasets are strongly correlated (0.73 for the number of mass graves and 0.65 for the number of victims in mass graves). However, in Galicia the correlation is around 0.10 and even stops being significant in the case of victims of mass graves, which is striking given that Galicia fell under Francoist control from the very beginning and virtually all violence was committed by the Francoist side. Again, in most cases, the

³See the Appendix for descriptives in linear scale.

Table 1: Correlation between Francoist killings and mass grave data, by region

Region	Mass graves (log)	Victims in MG (log)
All	0.37***	0.34***
Galicia	0.14*	0.1
Asturias	0.73***	0.65***
Basque Country	0.43***	0.47***
Catalonia	0.24***	0.21***
Aragon	0.43***	0.47***
Albacete	0.59***	0.38***
Badajoz	0.34***	0.25**

Note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Pearson correlation (ρ) between mass grave variables and the logged number of killings by Francoist forces between 1936 and 1942 in each region, according to the region-specific historical projects, in each municipality.

correlation with the number of mass graves is higher than in the case of victims of mass graves, which again points to another source of bias because of the non-exhumed graves and different factors that could influence the local number of victims in mass graves.

A clear difference between regions is that violence intensity over time varies greatly depending on the evolution of the conflict in each of them. During the Spanish Civil War, Francoist victimization took place predominantly after territorial conquest. Moreover, because violence dynamics did not remain constant throughout the conflict and the postwar period, the correlations could also be different depending on the date of the killing. Table 2 shows the correlation between the mass grave data and the victimization datasets distinguishing killings by year, for those regions where there is data on the date of death. The results show that the mass grave data does a particularly bad job registering killings during the first year of the conflict, while the correlation is higher in 1937 and 1938. After 1939, when the conflict ended, the correlation is again lower. Besides regional dynamics, these results could indicate that the use of mass graves was different as the conflict evolved. While killings during 1936, right after the outbreak of the conflict, do not match those registered in the mass graves data, the correspondence improves in the subsequent two years. After the conflict ends, the correlations between the two datasets decreases

Table 2: Correlation between Francoist killings and mass grave data, by period

Year	Mass graves (log)	Victims in MG (log)
All (1936-1942)	0.41***	0.39***
Killed in 1936	0.1**	0.16***
Killed in 1937	0.54***	0.42***
Killed in 1938	0.54***	0.45***
1939 or later	0.35***	0.35***

Note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Pearson correlation (ρ) between mass grave variables and the logged number of killings by Francoist forces in each period, according to the region-specific historical projects, in each municipality. The mass grave data does not discriminate by year of killing. Data is limited to those regions where there is data on the specific date of killings: Galicia, Asturias, Basque Country, Albacete, and Badajoz.

again, which is likely related to more systematic forms of carrying out repression in the postwar period, such as martial courts.

Another possible bias is that the temporal differences in the correspondence between the two datasets could be due to the type of killings. Table 3 shows these correlations distinguishing between *paseos* (strolls, or irregular killings usually committed by local militias) and executions (killings carried out in martial trials, for those regions where there is available data). A key difference between these two types of violence is that victims of *paseos* should, in principle, be much more likely to be buried close to their home. People who were sentenced to death could have been killed and buried in a place different from their residence. The data supports this point, at least for Asturias, the best documented region. The correlation of mass graves in each municipality with data on individual killings is higher for those victims who were killed irregularly, and lower for those executed. The correlations in Galicia and Badajoz, if much worse, point to a similar pattern. Interestingly, in every case the correlation is again lower for the number of victims in mass graves than for the number of local mass graves.

Finally, the data from Asturias allows for a fine-grained disaggregation of wartime victims. Table 4 shows the correlation between the mass grave data and the number of local victims, grouping together all Francoist killings, all Republican killings, all direct killings

Table 3: Correlation between Francoist killings and mass grave data, by type

Year	Mass graves (log)	Victims in MG (log)
Executions (Galicia)	0.08	0.08
Paseos (Galicia)	0.16**	0.1
Executions (Asturias)	0.44***	0.57***
Paseos (Asturias)	0.74***	0.6***
Executions (Badajoz)	0.31***	0.15
Paseos (Badajoz)	0.31***	0.24**

Note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Pearson correlation (ρ) between mass grave variables and the logged number of killings by Francoist forces between 1936 and 1942 by type and region, according to the region-specific historical projects, in each municipality. The mass grave data does not discriminate by year of killing. Data is limited to those regions where there is data on killing form: Galicia, Asturias, and Badajoz.

Table 4: Correlation between victims and mass graves in Asturias, by type and actor

Year	Mass graves (log)	Victims in MG (log)
All victims	0.66***	0.62***
Francoist victims	0.75***	0.66***
Republican victims	0.58***	0.53***
Direct killings (exec + paseos)	0.72***	0.63***
Combat victims	0.56***	0.59***
Non-categorized victims	0.47***	0.5***

Note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Pearson correlation (ρ) between mass grave variables and the logged number of killings in Asturias by type and actor, for the whole period. The mass grave data does not discriminate by year of killing.

(*paseos* and executions), combat victims, and non-categorized victims.⁴ Perhaps not surprisingly, the mass grave data is more strongly correlated with both Francoist victims and all victims of direct killings (including those committed by either side). The correlation is higher for the number of local mass graves than for the number of victims found in those groups, suggesting again that the distribution on the number of victims in each mass grave is not independent from the dynamics of violence.

⁴I show in the appendix linear models including these variables along with controls.

4 Accounting for spatial dependence

Spatial dependence is likely to be present in both these datasets. Besides spatially correlated errors which could influence the estimates, the relationship between wartime violence and the location of mass graves might be the product of non-local effects of predictors. Namely, if a given municipality was more violent during the war, neighboring municipalities could be more likely to host mass graves, as the perpetrators could bury the victims in the area.

Here I introduce spatial models to assess the relationship between these two datasets accounting for these dependencies, which should be taken into account when interpreting the previous correlations. In particular, in I compare a normal linear model with both a spatial error model (SEM) and spatial Durbin model (SDM).

Table 5 shows the results of these three models on both the number of local mass graves and the number of victims found in local mass graves. Overall, the results suggest the existence of spatial dependence. The λ coefficient ranges from 0.3 to 0.5 and it is statistically significant in all spatial models. The coefficient for the number of Francoist killings decreases, even if it remains significant, in the spatial models compared to the linear model. Interestingly, in the SDM models, the spatial lag of Francoist killings is also statistically significant and shows a meaningful effect size, pointing to non-local effects.

Finally, another finding in the results points to a potential bias in the use of proxies for the use of violence. In the SDM models, while the coefficient for population is positive and significant, its spatial lag is negative and significant, showing that mass graves were disproportionately located in more populated municipalities. Again, this result suggests that using a proxy variable such as the location of mass graves would involve attributing more deaths to larger municipalities, some of which correspond to neighboring, less populated towns. This is particularly worrying because it would introduce a systematic bias between urban and rural areas. Indeed, previous critiques to the literature on civil war research has already warned that relying on secondary historiography can introduce an ‘urban bias’ because of the source of the historical materials ([Kalyvas, 2004](#)). Even though

Table 5: Wartime Francoist victimization and log victims in mass graves

	Mass graves (log)			Victims in mass graves (log)		
	LM (1)	SEM (2)	SDM (3)	LM (4)	SEM (5)	SDM (6)
(Intercept)	-0.132* (0.067)	-0.618*** (0.074)	0.121 (0.132)	-0.205 (0.150)	-0.916*** (0.167)	0.740** (0.251)
Francoist killings	0.130*** (0.009)	0.097*** (0.009)	0.095*** (0.009)	0.267*** (0.020)	0.222*** (0.022)	0.194*** (0.022)
Elevation SD	0.001*** (0.000)	0.001*** (0.000)	0.000** (0.000)	0.001*** (0.000)	0.001* (0.000)	0.000 (0.000)
Log Population 1930	0.023* (0.010)	0.094*** (0.011)	0.103*** (0.011)	0.031 (0.022)	0.132*** (0.025)	0.219*** (0.027)
(Lag) Franc killings			0.043** (0.015)			0.121*** (0.032)
(Lag) Elev SD			0.000 (0.000)			0.001 (0.001)
(Lag) Log Pop 1930			-0.116*** (0.018)			-0.324*** (0.039)
λ		0.50***	0.48***		0.32***	0.30***
Observations	2,506	2,506	2,506	2,506	2,506	2,506
Adjusted R ²	0.158			0.120		
Log Likelihood		-1,294.292	-1,272.807		-3,478.963	-3,442.211

Note: + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

the bias mentioned here with respect to the data sources does not necessarily correspond to the data collection process, it would aggravate this problem.

5 Conclusion

This research note compares two previously used datasets on wartime victimization during the Spanish Civil War: regional lists of civilians killed and the location of mass graves. Results show that using the location of mass graves as a proxy for local wartime victimization would introduce systematic biases in the analyses. The relationship between these two data sources varies depending on the region, time, and the type of violence. Moreover, spatial analyses suggest that mass graves would be disproportionately located in more populated municipalities. In brief, the results show that these two variables correspond to a different data generating process.

Even if the analyses are limited to the Spanish Civil War, the goal of this note is to show, using of the few historical cases where we have alternative sources of data on violence, that indirect measures of violence need to be carefully assessed. The use of proxies introduces potential biases in the analyses because different types of violence leave different trails. Not accounting for this biases can invalidate empirical inferences.

In the last few years, empirical conflict research—and particularly the study of historical conflicts—has greatly benefited from a thorough attention to causal identification. However, methodological discussions on the problem of measurement are usually much shorter. This research note tries to highlight that measurement validity should also be taken into account.

References

- Armstrong DA, Davenport C and Stam A (2020) Casualty estimates in the rwandan genocide. *Journal of genocide research* 22(1): 104–111.
- Balcells L (2017) *Rivalry and Revenge: The Politics of Violence during Civil War*. New York: Cambridge University Press.
- Balcells L and Stanton JA (2021) Violence against civilians during armed conflict: Moving beyond the macro-and micro-level divide. *Annual Review of Political Science* 24.
- Balcells L and Sullivan CM (2018) New findings from conflict archives: An introduction and methodological framework. *Journal of Peace Research* 55(2): 137–146.
- Barceló J (2018) Batons and ballots: The effectiveness of state violence in fighting against Catalan separatism. *Research & Politics* 5(2): 1–9.
- Baum MA and Zhukov YM (2015) Filtering revolution: Reporting bias in international newspaper coverage of the Libyan civil war. *Journal of Peace Research* 52(3): 384–400.
- Douglass RW and Harkness KA (2018) Measuring the landscape of civil war: Evaluating geographic coding decisions with historic data from the Mau Mau rebellion. *Journal of Peace Research* 55(2): 190–205.
- García C, Alonso PL, Álvarez G, Cabrero C, Caunedo A and Díaz I (2011) *Base de Datos de Las Víctimas de La Guerra Civil y de La Represión Franquista En Asturias*.
- Hammond J and Weidmann NB (2014) Using machine-coded event data for the micro-level study of political violence. *Research & Politics* 1(2): 2053168014539924.
- Kalyvas SN (2004) The urban bias in research on civil wars. *Security Studies* 13(3): 160–190.
- Kalyvas SN (2008) Promises and pitfalls of an emerging research program: The microdynamics of civil war. In: Kalyvas SN, Shapiro I and Masoud T (eds.), *Order, Conflict, and Violence*, Cambridge: Cambridge University Press. pp. 397–421.
- Kasara K (2016) Electoral geography and conflict: Examining the redistricting through violence in Kenya. *Unpublished manuscript*. .

Kocher MA and Monteiro NP (2016) Lines of Demarcation: Causation, Design-Based Inference, and Historical Research. *Perspectives on Politics* 14(4): 952–975.

Krüger J, Ball P, Price M and Green AH (2013) It doesn't add up. methodological and policy implications of conflicting casualty data. *Counting Civilian Casualties: An Introduction to Recording and Estimating Nonmilitary Deaths in Conflict* : 247–264.

Tur-Prats A and Valencia Caicedo F (2020) The long shadow of the spanish civil war .

Walden J and Zhukov YM (2020) Historical legacies of political violence. In: *Oxford Research Encyclopedia of Politics*.

Weidmann NB (2016) A closer look at reporting bias in conflict event data. *American Journal of Political Science* 60(1): 206–218.