Unobserved Union Violence: Statistical Estimates of the Total Number of Trade Unionists Killed in Colombia, 1999-2008

Daniel Guzmán, Tamy Guberek, and Megan Price
Benetech Human Rights Program

November 13, 2012

Contents

1 Introduction

Anti-union assassinations may be one of the best documented violations in Colombia. Unions keep track of who has been killed in their community; companies and public institutions know when their employees are no longer alive. Special interest groups have formed to pay even closer attention to violence across unions, labor sectors, years and municipalities. The National Union School (*Escuela Nacional Sindical* in Spanish) is one such group: they have recorded violence against trade unions since 1986. Human rights groups, such as the non-governmental (NGO) Colombian Commission of Jurists, monitor the killing of unionists in addition to other victims and other violations. The government's human rights office, led by the Vice President, began pooling together data on Colombia's anti-union violence in 1999. Yet, even with so many monitors keeping track of the violence experienced by this group, we estimate in this study that in some places and times, as much as 30% of all the killings are not recorded by any of these datasets.

Many have attempted to answer the apparently simple question: "how many trade unionists have been killed in Colombia?" Unions, human rights groups and the Colombian government debate the "right" answer. More recently, the political stakes of the answer increased when governments in North America and Europe linked their consideration of free trade agreements with Colombia to the question of whether violence against unionists has been improving or not. As part of the answer, some groups suggest lists of killings. For example, the ENS published a report in 2007 entitled *"2.515 o esa siniestra facilidad para olvidar,"* emphasizing their recorded death toll through that date [Montoya, 2007]. Their most recent published count is 2,857 union workers killed in the 25 years between 1986-2010 [Escuela Nacional Sindical, 2011]; a recent news article presents a different number, 2,819 unionized workers killed, for the same period [El Espectador, 2011]. Beyond counts, others try to suggest a pattern across time. For example, the UN High Commissioner for Human Rights cites 26 union assassinations in 2010 as an increase from the 25 reported in 2009. The ENS disputes those yearly numbers with a higher count on their website - 28 assassinations in 2009 and 51 in 2010.[1] The controversy around this apparently simple question suggests that there may be more killings that are unknown to some of the monitoring organizations. The debate becomes more complicated when researchers attempt to provide explanations about why the killings occur (or why they do not occur), occasionally with politically-charged interpretations. In 2009, for example, a study argued that union violence in Colombia is neither systematic nor targeted, an argument that many have disputed [Mejía and Uribe, 2009].[2] The Achilles' heel of all these claims is that the data are incomplete. All of the existing datasets document some killings, but none of the databases include all trade union homicides.

Faced with tremendous political pressure to provide statistics – and to defend them – analysts tend to explain differences between their figures and other analysts' counts as follows. Some analysts claim that competing numbers are inflated because they include deaths of trade unionists killed for non-union-related motives. Other analysts claim that the rival numbers are deflated because they categorically exclude some of the victims. The datasets used in this study included cases of killings against unionists related to their union activity - organizations that contributed data to this study carefully selected cases of killings of unionists which were viewed to be directly related to their union activity, other kinds of killings such as common crime or crimes of passion are not included. However, even with this common definition of what constitutes a 'case' we find some disagreement between groups about whether or not a homicide is related to union activity. As a result, we conducted analyses examining the sensitivity of the results and conclusions presented here to differing decisions regarding the inclusion or exclusion of homicides in the datasets used in this report. For more details see Appendices B and C. Although there is general agreement that even with a particular definition, none of the databases includes *all* of the victims, researchers have yet to sufficiently address this limitation.

We have two purposes with this study. The first is to offer statistical estimates of unknown, unreported trade unionists killed. We use a statistical method called Multiple Systems Estimation, explained in detail in Appendix A. By combining an estimate of the number of deaths not recorded in any source with the number of known, recorded deaths, this analysis offers estimated totals for all unionists killed in Colombia between 1999 and 2008.

[1] http://www.ens.org.co/index.shtml?apc=a---;1;-;-;&x=20166326

[2] For a critique of the study, see Price and Guzmán [2010]

Second, we are concerned that underregistration (deaths not recorded by any source) has not been sufficiently understood in previous studies which aim to measure anti-union violence[3], human rights violations[4] or war deaths[5]. In this report, we demonstrate the interpretative problems that result from ignoring underregistration. We show that underregistration of union killings is not constant across the dimensions we are trying to understand including time, space, and type of union. The variation in levels of undercounting is important because it can alter the interpretations we make when comparing across these dimensions. We discuss our high degree of confidence in the models that are correcting the underregistration biases in Section 4 and in Appendix A.

There are three important datasets that record violence against union members: the National Union School (ENS), the Vice Presidency's Human Rights Observatory (VP), and the Colombian Commission of Jurists (CCJ); each contributes an important, but partial, piece of the picture of union homicides in Colombia. We use statistical modeling to combine and interpret these datasets to tell a more complete story of union homicides. Before presenting our estimates of the total number of union killings and the patterns observed across the estimates (Section 3), we will first describe and discuss the individual datasets (Section 2). We believe it is important for the public interested in measurements of violence to understand what is behind the various statistics suggested by these datasets. In Section 4 we discuss the relevance of the findings. In Section 5, we present conclusions of this study.

In their recent book, Andreas and Greenhill suggest that "...given the chronic and pervasive nature of political use and abuse of numbers, it behooves consumers of numbers to assess them with a critical eye and ask hard questions about their origins, even if doing so requires consumers to step outside their numeracy comfort zones" [Andreas and Greenhill, 2010, 3]. We hope that with a careful explanation of the data and methods behind any quantified claim about violence, the policymakers –that base decisions on these claims– can be better consumers of the figures. Calculating an accurate estimate requires advanced statistical methodology. Consumers of the numbers should not accept partial answers or interpretations. None of the existing datasets are wrong, they are just incomplete in ways that only become clear when multiple datasets are combined and analyzed to estimate the undocumented cases.

The people with the highest stakes in a more accurate estimate about the patterns and magnitude of the violence are the victims themselves. Getting the numbers right can help account for unnamed, unreported victims in the historical record and guide the development of policies to respond to past violence. Using biased or incomplete figures, on the other hand, risks losing all trace of the existence of some victims. Victims who remain undocumented by any dataset become invisible, removed not only from their lives and the lives of their families, but from historical memory.

With these estimates, we hope to use statistical analysis to transcend a highly politicized and important debate, thus increasing the possibility of developing proper policies to prevent future atrocities.

2 Exploring Observed Patterns

Each monitoring organization on union violence records a different set of killings. Naturally, each organization can better observe some cases over others. For example, a union located in Medellín, the capital of Antioquia, will likely know more about violence against unionists in Antioquia than they would know about violence in Arauca, several departments (or states) away. A government agency may know more about public sector killings than about killings in the agricultural sector. However, when the time comes to explain different numbers, these epistemological explanations about the social process of collecting data are rarely offered.

By emphasizing the importance of underregistration, we are arguing that the data recorded in each database are not by themselves sufficient for a meaningful quantitative understanding of the magnitude and patterns of killings against

[3] Mejía and Uribe [2009], Montoya [2007], Sanjuán et al. [2010], Observatorio del Programa Presidencial de Derechos Humanos y DIH [2009] and U.S. Labor Education in the Americas Project (USLEAP) [2009]

[4] Clark and Sikkink [2011]; Dube and Naidu [2010]; Hicks and Spagat [2008]

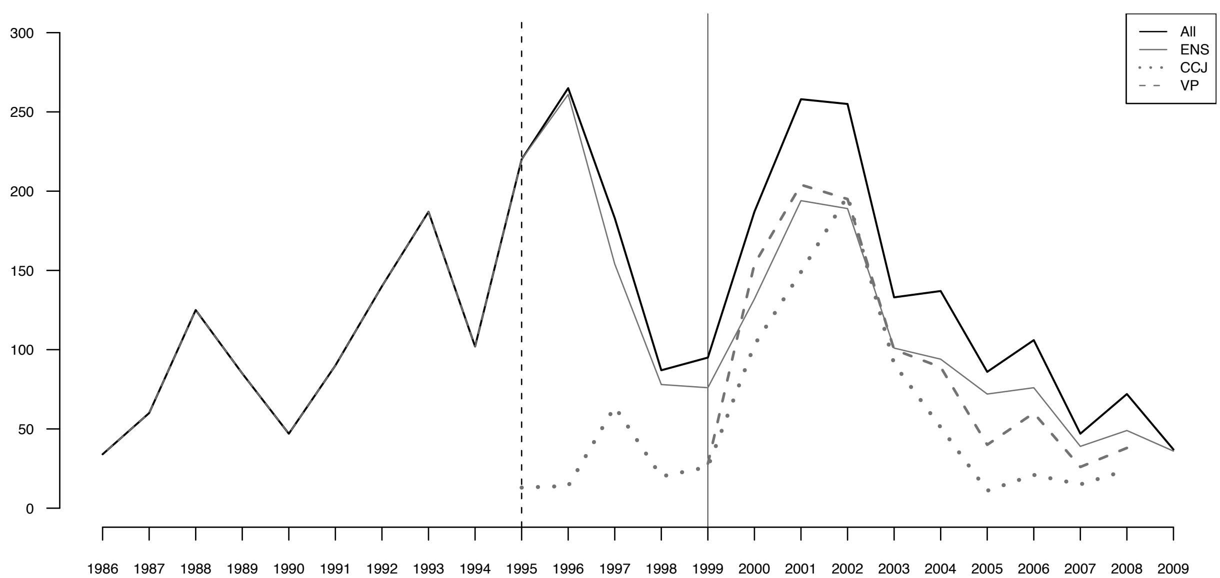
[5] Bohannon [2011]; Lacina et al. [2006]

unionists. This is because the majority of existing sources of data on union violence, and specifically the three datasets examined in this report, are *convenience samples*. Convenience samples are datasets collected without an underlying probabilistic selection method (i.e., by random sampling). This means that existing sources are not necessarily representative of the entire population of victims of union violence. Indeed, *convenience samples* correlate with the underlying patterns in the population only by coincidence. Such datasets may include an *unknown and biased* subset of the underlying population for a variety of reasons - access to certain sectors of the population, resource limitations, security concerns, etc. See Guberek et al. [2010] and Gohdes [2010] for more detailed examples of reasons why and ways in which convenience samples can be incomplete and contradictory.

2.1 Observed Patterns Over Time

Figure 1 plots on one graph the killings recorded by each organization by year. This kind of comparison of datasets has been published in multiple reports. The three organizations themselves have published their single source in their own reports. All of the reports highlight a peak in reported killings in 1996 (exclusively based on the ENS database) and another in 2001 (observed in all three datasets).

Figure 1: Observed Killings by Year



In addition to overlaying the data from our three sources in one figure (the lighter grey lines labeled ENS, CCJ, and VP in Figure 1), we have added a solid black line labeled 'All.' This line represents an aggregation of all three datasets. More specifically, using the "raw data" (including the names, dates and places of each death), we have "matched" all three datasets. Matching means de-duplicating names within lists (i.e., identifying multiple records *within the same list* that refer to the same person) and linking names across lists (i.e., identifying multiple records *across different lists* that refer to the same person). See Appendix D for details on how records were matched in this study.

The solid black line in Figure 1 indicates the total number of reported cases with the three datasets matched and merged, thereby counting each victim once. The dotted vertical line indicates when the CCJ started their database (in

1995), and the solid vertical line indicates when the VP started their database (in 1999). Since 1999 is the year union violence is observed by all three datasets, it is the starting point for the estimates in this study.

The three sources have the same reporting pattern after 1999, but notice the great differences in reporting before 1999. In the period 1996-1999, the CCJ and the ENS would tell very different stories. For example, if we relied solely on data from the ENS, we would conclude that following a peak in 1996, union violence steadily decreased until 1998 when it leveled off or potentially increased slightly over the following year. In contrast, if we only had access to the CCJ data we would conclude that union violence was at an all-time low in 1996, peaked in 1997, then decreased the following year before leveling off. Not only would we describe the pattern of violence differently depending on which dataset we used, but the magnitude of violence would differ substantially by dataset - peak violence in the CCJ dataset registers slightly more than 50 killings whereas the peak in the ENS dataset records more than 250 killings. Unfortunately we can only speculate about pattern and magnitude of union violence during these years, since it is not possible to calculate multiple systems estimates until 1999 when all three data sources begin recording cases.[6]

It is important to note that after 1999, although all datasets and the matched data (indicated by the solid black line) follow a similar pattern, the matched data consistently indicates a higher number of killings than recorded in any individual dataset. This implies that none of the datasets are subsets of each other. For example, in 2001 the VP records a higher number of killings than either the ENS or CCJ. So one might argue that the VP data in 2001 is “complete” and ENS and CCJ are both missing records. But the matched data records an even higher number of killings in 2001, indicating that there must be records in VP that are not in CCJ or ENS but also that there must be records in CCJ or ENS that are not in the VP dataset. While all datasets are partial, each reports some cases that are not in either of the other two. In other words, each dataset adds knew knowledge. This is one example showing why no one dataset alone should be used to explain violence.

As a thought experiment, how many hypothetical new datasets would have to be added before no new victims were found? Asked differently, how many more victims are not recorded by any of the three datasets? It is impossible to know the full picture of union violence until the underregistration can be estimated.

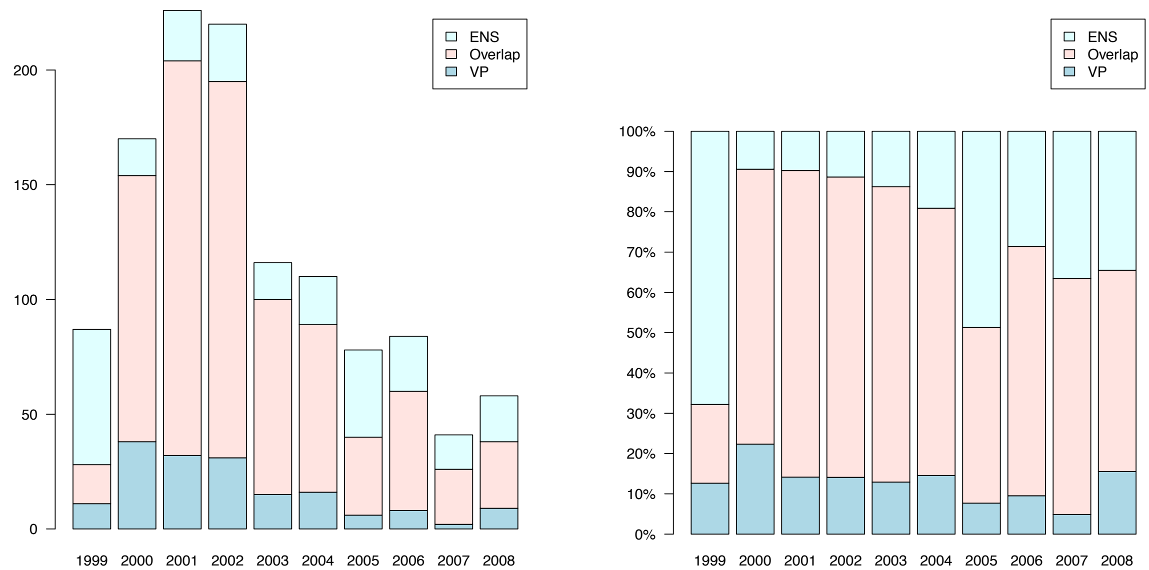
2.2 Records in More Than One Dataset

Another way to visualize differences in datasets is by looking more closely at two datasets at a time. Figure 2 shows the amount of overlap between the ENS and VP datasets. Overlap refers to the collection of records that appear in both datasets. The dark blue at the bottom of the bar chart represents the total number of records (a) or proportion of records (b) that appear only in VP and not in ENS. The pink in the middle of the bar chart represents the number of records (a) or proportion of records (b) that appear in both VP and ENS. Lastly, the light blue at the top of the bar chart represents the number of records (a) or proportion of records (b) that appear only in ENS and not in VP.

Figure 2 shows that the VP and ENS datasets have very little in common in 1999, which is not surprising given the pattern in 1999 in Figure 1. The following years, until 2003, indicate a high degree of overlap between the two datasets, which aligns with the contextual information that the ENS and VP collaborated during this time.

[6]Technically, a two-systems estimate could be calculated for the years prior to 1999 using data from ENS and CCJ. However, we expect the assumptions required to calculate estimates from two-systems to be violated, and therefore, we only calculate estimates after 1999 when when we have access to three datasets - see Appendix A for more details.

Figure 2: Overlap Counts and Rates for ENS and VP



(a) Counts (b) Rates

Visualizing the overlap between all three datasets at once requires slightly different displays. These are presented as venn diagrams in Figure 8 and will be described more fully in Section 3.1.

2.3 Observed Patterns Over Space

The bar graphs below show observed patterns by department, looking specifically at the subset of years covered by all three datasets (1999-2008). Figure 3 uses the matched dataset (each repeated victim only counted once) and displays the observed number of killings over space whereas Figure 4 displays these same observations for each individual dataset.

It is important to note that the patterns in Figures 3 and 4 are not identical. The most notable cases are Putumayo, Bolívar, Meta and Santander. Graphs 3 and 4 highlight these departments to illustrate how the pattern changes depending on which list is chosen. For example, the VP reports fewer numbers of killings in Bolívar than in Tolima and fewer in Meta than in Cauca. The ENS reports fewer killings in Meta than in Bolívar, while the CCJ reports exactly the opposite - fewer killings in Bolívar than Meta. The CCJ also reports more killings in Putumayo than Bolívar. This means that if we were to rely on any single dataset to report the relative severity of violence over geographic regions, perhaps for the purpose of allocating resources, we would draw different conclusions depending on which dataset we used.

It is also important to note that this is not the best way to observe relative severity. Comparisons of counts indicate potentially interesting patterns, but we must also consider the population of each department. In Section 3, we will look at the estimates by department as population adjusted rates.

Figure 3: Total Observed Killings, by Department, Between 1999 to 2008

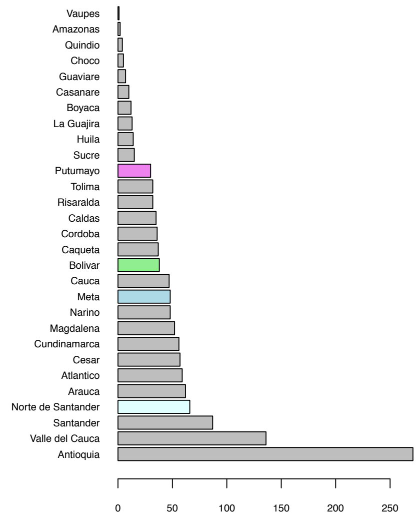
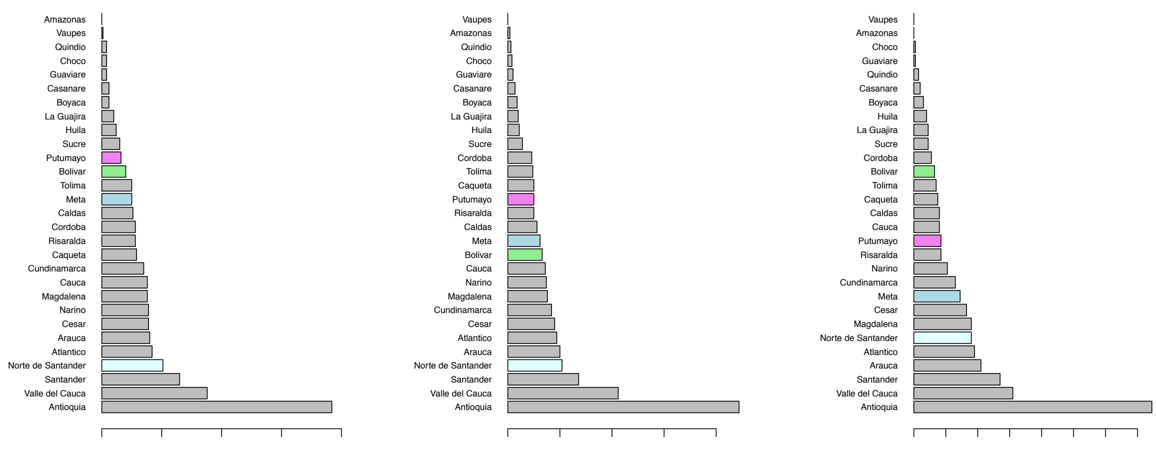


Figure 4: Total Observed Killings, by Department and Dataset, Between 1999 to 2008



(a) VP (b) ENS (c) CCJ

2.4 Observed Patterns by Union Sector

Similar to the comparisons in the previous section, Figure 5 shows the observed number of killings across union sector in the matched dataset and Figure 6 shows these same observations for each individual dataset. In contrast to the previous graphs of union violence over geographic area, the patterns reflected across union sector in the individual datasets (Figure 6) much more closely match the pattern in the matched dataset (Figure 5). However there is one notable exception - CCJ reports fewer killings in the judicial sector than in 'other' (and several preceding sectors) whereas ENS, VP, and the matched dataset all indicate an increase in the number of killings in the judicial sector as compared to 'other' (and preceding sectors). Again, this could lead to misallocation of resources if policy decisions were based on the relative severity of violence across sectors using only one dataset.

Also, as with comparisons of observed counts by department, we must keep in mind the number of individuals in each union sector. The patterns in Figures 5 and 6 raise interesting questions, but we must also look at the distribution of the *proportion* of members of each sector killed to gain the full picture of union violence.

Figure 5: Total Observed Killings, by Union Sector, Between 1999 to 2008

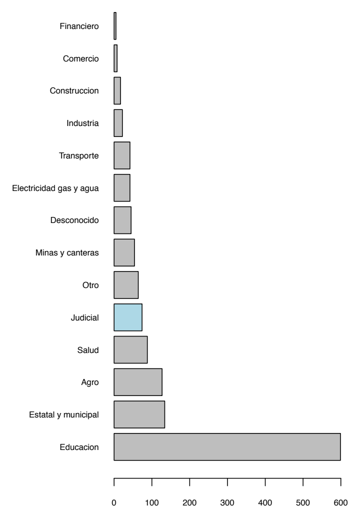
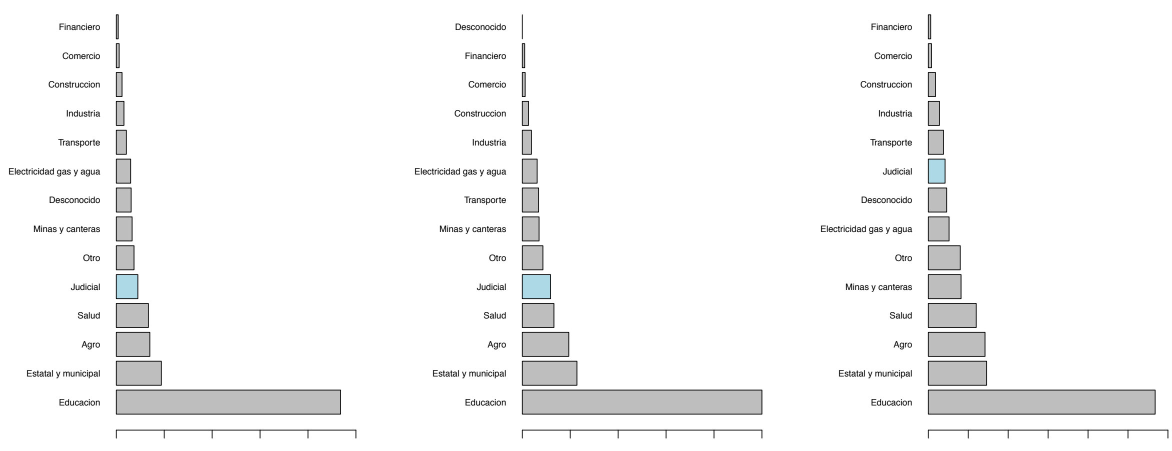


Figure 6: Total Observed Killings, by Union Sector and Dataset, Between 1999 to 2008



(a) VP (b) ENS (c) CCJ

Lastly, it is worth noting that we have presented observed patterns, and in the following sections will present calculated estimates, based on variables that were available in all three datasets. There are any number of other metrics on which we would like to compare union violence, such as perpetrator, or municipality instead of department. Unfortunately data limitations prevent us from calculating estimates beyond the ones presented in the following sections.

3 Estimates of the Magnitude and Patterns of Trade Unionists Killed, 1999-2008

As we have seen in Sections 1 and 2, we cannot assume that any single dataset provides a complete or representative picture of the true underlying population of all union killings. Each data source includes some, but not all killings. In addition to matching records across data sources to generate a single dataset that includes all observed killings, we must also estimate the number of *unobserved* killings. We use a statistical method called multiple systems estimation (MSE) to estimate the number of killings not recorded by any of the data sources.

MSE calculations rely on information about the size of the 'overlap' of records between datasets, as described in Section 2.2 and illustrated in Figure 2 by two sources and in Figure 8 with *three* sources. By analyzing the pattern of these overlapping records, we are able to estimate statistically how many victims are not reported at all, and therefore the final estimates include both observed and *unobserved* killings. The following sections present results based on MSE analyses; a technical introduction to MSE as well as the specific method used to calculate the results in this report can be found in Appendix A.

It should be noted that all of the estimates presented in this report are stratified - they are estimates for individual years or departments or union sectors. Due to the structure of the data and the assumptions of MSE we need to stratify to these smaller units to calculate more reliable estimates (see Appendix A for more detail). For these reasons we do not present an overall total estimate of the number of union members killed across all of Colombia for all ten years analyzed in this report. We understand the impulse to sum estimates across strata to approximate such a total for the entire country and time period under study; indeed we present such sums in Appendix A. Although these sums are all quite similar, it is important to note that they are not expected to be identical, since each stratification models the data slightly differently and accounts for slightly different sources of variation.

3.1 Estimates by Year

Figure 7 shows estimates calculated by year for the period 1999-2008. One important requirement of the MSE model we used is that at least three data sources must contribute to each estimate (see Appendix A for an explanation of this requirement). As a consequence, we were not able to make estimates for some of the most violent years of anti-union killings.

The bar graphs below present the point estimate and associated error bars for each year. The lower part of each bar in the graphs represents the number of victims reported in the matched dataset for each year. Stacked on each bar is the estimated quantity of underregistration. The top of each bar, therefore, indicates the point estimate for the total number of victims. To be clear, the amount of underregistration refers to the number of killings not recorded by any of the three datasets (ENS, VP and CCJ). The lines, or "whiskers" around each bar indicate the 95% confidence interval for each specific estimate. We also present all the numbers plotted on the graph in tables next to each graph.

Among the most salient findings is that in 2001, the year with the most reported killings, we estimate the largest proportion of underregistration. While the number of reported killings in 2001 was 258, the estimated total number of killings (observed and unobserved) is expected to be between 308 and 620, and point estimate of 386.[7] Looking at the estimates by year, the higher the number of reported killings, the greater the estimated underregistration rate. In other words, the estimated total and the rate of underregistration of killings are highly correlated over time (the correlation

[7]The reported range comes from the 95% confidence interval, and the point estimate is the maximum likelihood estimate. See Appendix A for a more complete explanation of these calculations and the methodology behind them.

coefficient is 0.74).[8]

Regarding patterns, we should note that 2001 and 2002 looked almost identical in the reported data. The estimates suggest that there were more victims in 2001 than in 2002. However, since the confidence intervals for these two years overlap, we cannot be certain in this conclusion.[9] Similarly, the number of victims in 2003 and 2004 looked almost identical in the reported data. The estimates suggest that there is a higher total for 2003 with 23% more killings than were observed, and with larger uncertainty. In this case, the confidence intervals just barely overlap, indicating that there may indeed be a meaningful difference in the total number of killings in 2003 compared to 2004.[10] It appears that since 2004, the 3 groups have covered the total universe more completely in their reports, although there still remains between 2% and 16% of possibly *unobserved* killings. There also appears to be a significant amount of underregistration in 2008, indicating an increase in the number of killings, a decrease in the proportion of killings that are observed, or, most likely, both. In general, underregistration decreases between 2004 and 2007.

Lastly, it is important to note that the pattern of underregistration is not constant across years. The size of the top section of each bar in Figure 7 varies by year. This can also be seen in the varying proportion of unobserved killings reported in the last column of Table 1. This means that the amount of unobserved killings changes from year to year, which makes it impossible to draw reliable conclusions about the pattern of violence over time based only on observed data.

Table 1: Estimates by Year

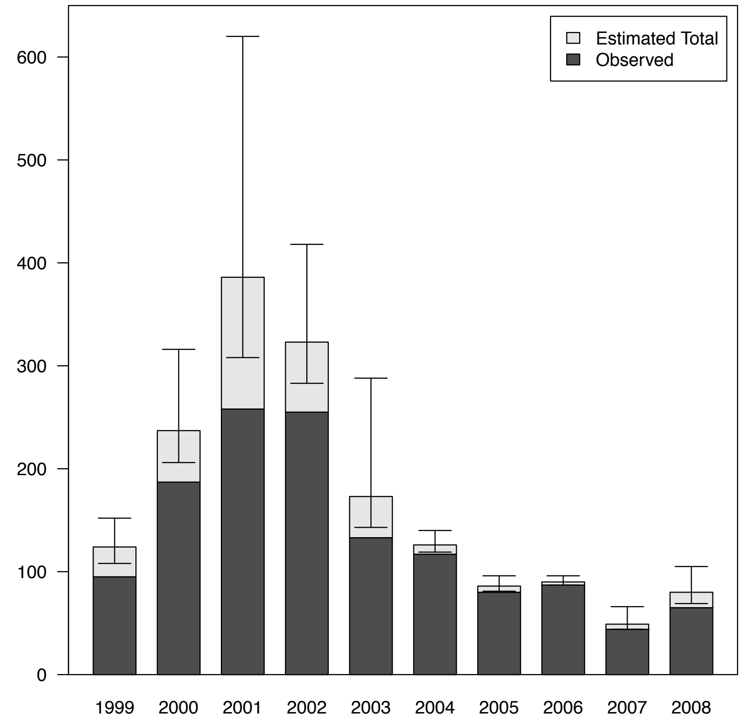
Year	Observed Killings	Estimated Killings	Confidence Interval 95%	Percentage of Underregistration
1999	95	124	(108 , 152)	23%
2000	187	237	(206 , 316)	21%
2001	258	386	(308 , 620)	33%
2002	255	323	(283 , 418)	21%
2003	133	173	(143 , 288)	23%
2004	117	126	(119 , 140)	7%
2005	80	86	(81 , 96)	7%
2006	87	90	(87 , 96)	3%
2007	44	49	(44 , 66)	10%
2008	65	80	(69 , 105)	19%

[8] We used Pearson's correlation coefficient, a statistical measure to quantify the strength of the linear relationship between two variables, in this case, underregistration and total magnitude.

[9] More formally, $H_0 : \widehat{N_{2001}} = \widehat{N_{2002}}$ vs. $H_1 : \widehat{N_{2001}} > \widehat{N_{2002}}$ P-value = 0.2022

[10] More formally, $H_0 : \widehat{N_{2003}} = \widehat{N_{2004}}$ vs. $H_1 : \widehat{N_{2003}} > \widehat{N_{2004}}$ P-value = 0.0637

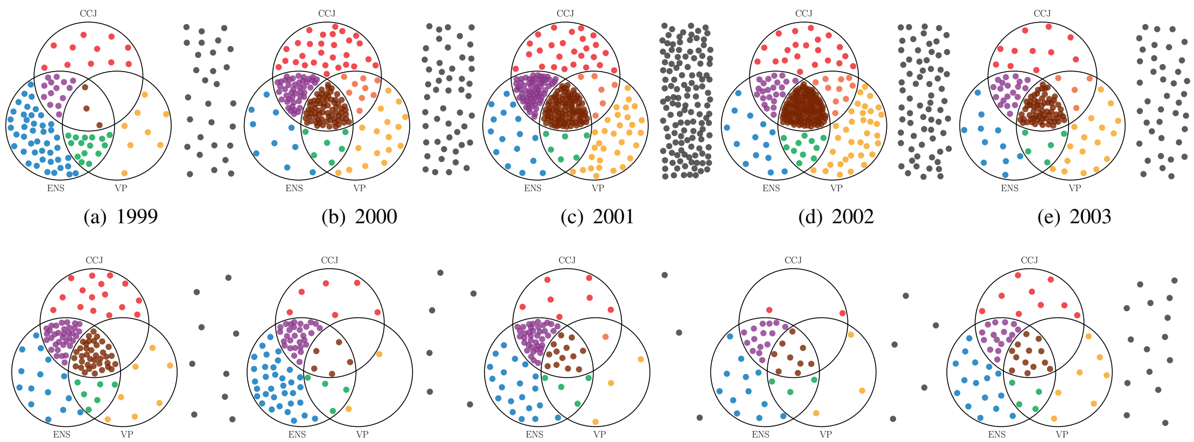
Figure 7: Estimated Union Killings by Year, 1999-2008



As mentioned in Section 2.2, one way to visualize the relationship between all three datasets is through a venn diagram. Figure 8 summarizes the overlap between all three data sources for the years 1999-2008. Each circle represents one data source (ENS, CCJ, VP), and the density of colored dots within a circle represent the number of observations recorded in that dataset. The colored dots in sections where two or three circles overlap represent the number of observations recorded in a pair or all three of the data sources. Again, the density of dots implies the relative number of records. Figure 8 includes an additional piece of information as well - the dots to the right of the three circles represent the estimated number of unobserved killings (i.e., killings not reported by any of the three datasets) as determined by MSE calculations.

For example, we can see that in 2001 there was a relatively high degree of overlap between all three data sources (the dense red color in the center where all three circles overlap) and between CCJ and ENS (the dense purple color where only two circles overlap). Yet there is also a large estimate of unobserved records, as indicated by the relatively dense rectangle of dots to the right of the circles. This also coincides with our interpretation of the bar graph in Figure 7.

Figure 8: Illustration of List Overlap and Point Estimates by Year.



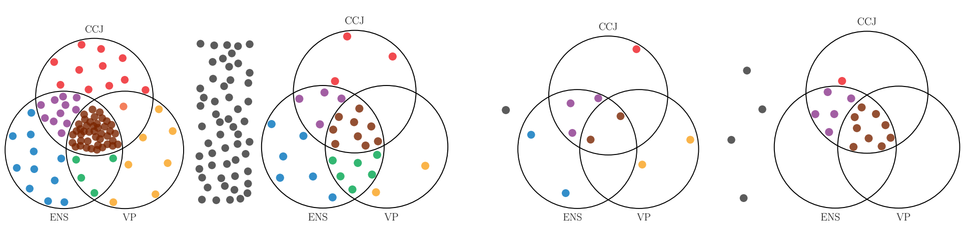
(a) 1999 (b) 2000 (c) 2001 (d) 2002 (e) 2003

(f) 2004 (g) 2005 (h) 2006 (i) 2007 (j) 2008

3.2 Estimates by Geographic Space

Estimates across departments (equivalent to U.S. States) indicate two important findings - (1.) although estimates in some departments indicate that the matched dataset is likely to include nearly all killings, there are other departments where unregistered homicide rates are high, between 25% and 38% (see Figure 10 and Table 2) and (2.) it is not possible to calculate estimates for a number of departments, due to the size and structure of the available data. The reason why estimates cannot be calculated in some departments will be elaborated shortly. First, we focus on departments for which it is unlikely that there are unobserved killings versus those for which estimates indicate a significant number of unobserved killings. This is determined by the 95% confidence interval - whether or not it overlaps with the observed number of records (the shaded section of the bar) indicates whether we can conclude with statistical significance that there were additional unobserved killings. Estimates and confidence intervals in Figure 10 indicate a statistically significant number of *unobserved* killings in Cordoba, Meta, Cundinarmarca, Santander, Valle Del Cauca, and Antioquia. however, it is also important to note that these estimates are accompanied by large ranges of uncertainty. In particular, there is a substantial underregistration of killings of unionists in Antioquia and Valle del Cauca.

Figure 9: Overlap Structure by Department, Between 1999 to 2008

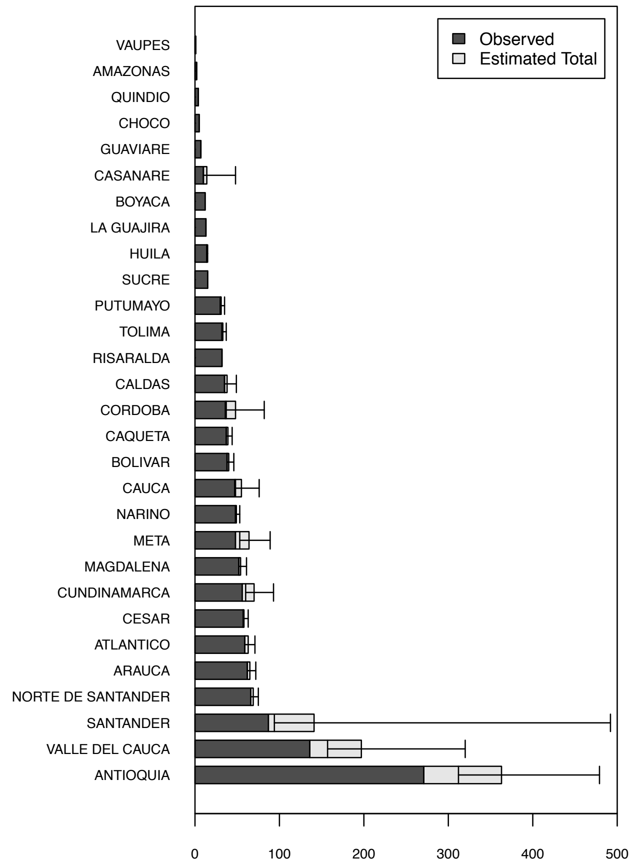


(a) Santander (b) Putumayo (c) Casanare (d) Sucre

Table 2: Estimates by Department

Department	Observed Killings	Estimated Killings	Confidence Interval 95%	Percentage of Underregistration
Vaupés	1	-	(- , -)	-
Amazonas	2	-	(- , -)	-
Quindío	4	-	(- , -)	-
Chocó	5	-	(- , -)	-
Guaviare	7	-	(- , -)	-
Casanare	10	14	(10 , 48)	29%
Boyacá	12	-	(- , -)	-
La Guajira	13	-	(- , -)	-
Huila	14	14	(14 , 15)	0%
Sucre	15	-	(- , -)	-
Putumayo	30	31	(30 , 35)	3%
Tolima	32	33	(32 , 37)	3%
Risaralda	32	-	(- , -)	-
Caldas	35	38	(35 , 49)	8%
Córdoba	36	48	(37 , 82)	25%
Caquetá	37	39	(37 , 44)	5%
Bolívar	38	40	(38 , 46)	5%
Cauca	47	55	(48 , 76)	15%
Nariño	48	49	(48 , 53)	2%
Meta	48	64	(53 , 89)	25%
Magdalena	52	54	(52 , 61)	4%
Cundinamarca	56	70	(60 , 93)	20%
Cesar	57	58	(57 , 63)	2%
Atlántico	59	63	(59 , 71)	6%
Arauca	62	65	(62 , 72)	5%
Norte de Santander	66	69	(66 , 75)	4%
Santander	87	141	(94 , 492)	38%
Valle del Cauca	136	197	(157 , 320)	31%
Antioquia	271	363	(312 , 479)	25%

Figure 10: Estimated Union Killings by Department, Between 1999 to 2008



The relatively small number of observed killings in several departments prevents the calculation of estimates in some departments and results in large amounts of uncertainty in others (see Table 2). Statistical estimates are generally more uncertain when they are based on less data. In other cases, such as the department of Santander, a relatively large amount of observed data still produces an estimate with a large amount of uncertainty. In this case the structure of the data - the pattern of overlapping records - leads to more uncertainty.

Both of these data challenges are shown in Figure 9, which presents the overlap structure for Santander, Putumayo, Casanare and Sucre. In departments where the overlap structure between lists is like the one in Sucre - nearly all observed killings recorded in two or all three of the datasets and very few or no observations unique to the individual datasets - we were unable to calculate estimates because there is not enough information to support the statistical model. However in departments like Putumayo or Casanare, with relatively small numbers of observations, the distribution of those observations among the possible combinations of lists was sufficient to estimate the underregistration. Lastly, we can see in Figure 10 and Table 2 that Santander has a relatively large number of observed killings and yet also one of the largest confidence intervals (indicating a high level of uncertainty). Figure 9 indicates the structure of the data that leads to this uncertainty in the estimate for Santander. The majority of observed records in Santander were recorded by all three datasets, which sometimes implies a smaller estimate of unobserved deaths - if all three datasets appear to be recording the same events, it is plausible that the underlying population has been sufficiently covered and very few events have been missed. However, in Santander, despite this relatively large number of records in all three datasets, there are 33 additional killings recorded only by VP, only by ENS, or by both VP and ENS. These remaining records pull the estimate up because they imply that there are areas of the underlying population that only one or two of the datasets are covering. And this skewed distribution of overlapping records increases the uncertainty in the estimate.

In general, large confidence intervals should motivate us to collect more data and to look more closely at the structure of the data to inform more targeted future data collection.

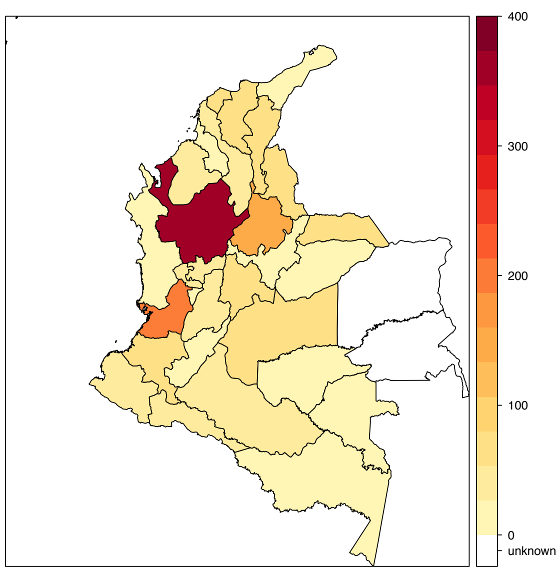
Another way to visualize estimates over space is by projecting them onto a map. Figure 11(a) illustrates the point estimates of unionists killed between 1999-2008 by department. It is important to keep in mind that the confidence intervals presented in Figure 10 and Table 2 are not included in Figure 11(a). As we saw in Figure 10, the estimates of total killings were highest for Antiqouia and Valle del Cauca. Santander is a special case, as the level of uncertainty is extremely high. The underregistration rate is most likely between 7% and 82%, with the point estimate being 38%.

As was noted in Section 2.2, departments are not properly comparable if we do not account for the different sizes of the unionized population in each. Therefore, we have used the census on unionized workers carried out by the ENS to offer an alternative perspective (Figure 11(b)).

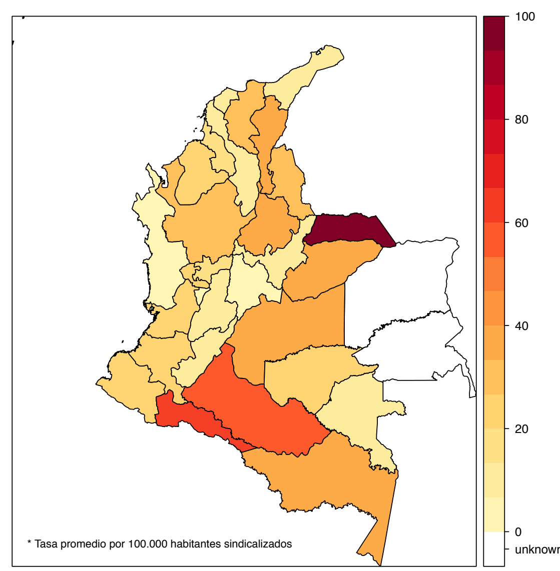
Usually homicide rates are presented by years, but given the available data, it was not possible to make estimates disaggregated by year and department at the same time. Therefore, we built the rate as follows: for the numerator, we used the average number of estimated killings per year for each department; for the denominator, we used the 2005 census of unionized workers carried out by the ENS.

Figure 11(b) shows a different pattern from Figure 11(a). The distribution of union workers killed between 1999-2008 is much more homogeneous than we would have noticed by only looking at the total by department. In other words, the proportion of union workers killed out of all the unionized workers appears more similar throughout the country than we could observe from estimates of totals. Furthermore, we are able to see that Arauca's unionized workers were proportionally the hardest hit in the country. Caquetá and Putumayo's trade unionists have also been killed in substantial proportions.

Figure 11: Estimates Killings Between 1999 to 2008. Total and Rates



(a) Point Estimates of Trade Unionists Killed by Department



(b) Rates of Estimated Killings by Unionized Population*

3.3 Estimates by Unionized Sector

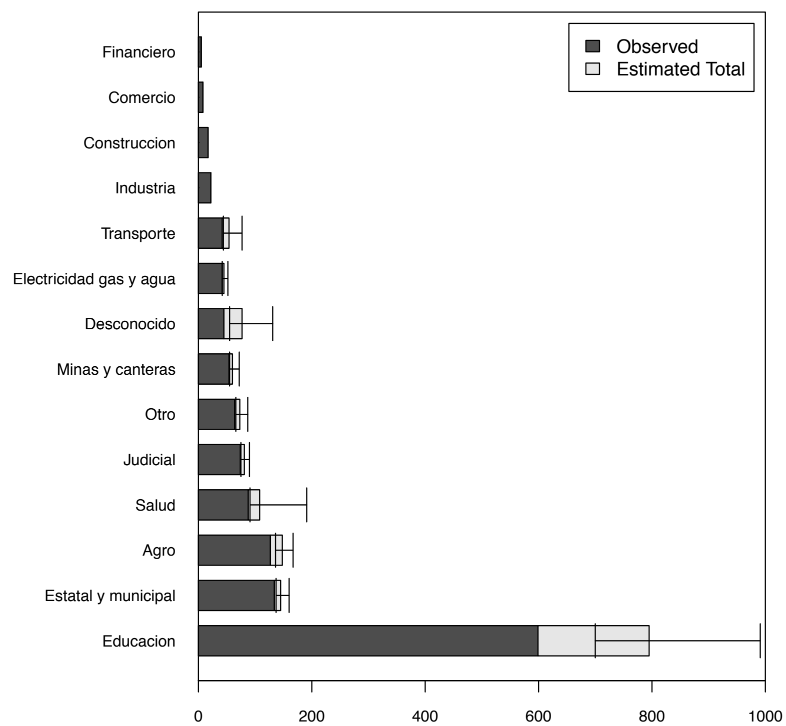
Much like the results presented in Section 3.2, Figure 12 indicates that all killings are recorded in the matched dataset for many of the union sectors. In contrast, both agriculture and education indicate a statistically significant number of *unobserved* killings. Are there important differences between these two sectors and all the others that may lead to less complete reporting of killings? Do more urban or higher paying sectors have better registration of violence?

As noted in Section 2, the highest number of recorded killings occur in the education sector. Yet despite this already large measured amount of violence, estimates indicate that an even higher number of killings are occurring, since a significant number are unobserved.

Table 3: Estimates by Union Sector

Union Sector	Observed Killings	Estimated Killings	Confidence Interval 95%	Percentage of Underregistration
Financiero	5	-	(- , -)	-
Comercio	8	-	(- , -)	-
Construcción	17	-	(- , -)	-
Industria	22	-	(- , -)	-
Electricidad gas y agua	42	45	(42 , 52)	7%
Transporte	42	54	(44 , 77)	22%
Minas y canteras	54	60	(55 , 72)	10%
Otro	64	73	(66 , 87)	12%
Desconocido	45	77	(55 , 131)	42%
Judicial	74	81	(75 , 90)	9%
Salud	88	108	(91 , 191)	19%
Estatal y municipal	134	145	(137 , 160)	8%
Agro	127	148	(136 , 167)	14%
Educación	599	795	(700 , 991)	25%

Figure 12: Estimated Killings by Union Sector, Between 1999 to 2008



4 Relevance of the Estimates

The ENS, the VP and the CCJ make enormous efforts and dedicate many resources to collecting data on union homicides. Our concerns about underregistration should in no way be interpreted as a critique of their work or of the quality of their data. Underregistration is an inevitable characteristic of any data collection process. Perpetrators may deliberately intend to hide violence. Varying degrees of access, conflicting interests, differing availability of resources, and many other aspects of data collection each affect what can be observed at all.

Therefore we must rely on statistical techniques to estimate the amount of underregistration and generate an estimate of the total number of killings. As described in Appendix A, the technique used in this report, multiple systems estimation, like all statistical techniques, relies on assumptions which must be evaluated within the context of the observed data. For example, we know from talking with the ENS and the VP that they collaborated and shared data particularly in 2001-2003. The collaboration was reflected in the structure of the models selected as the best fit (see Appendix A for modeling and model selection details). This relationship also appears in the overlap rates in Figure 2. During 2001-2003 we see large, relatively constant rates of overlap, and as of 2004 the overlaps change. This consistency between contextual knowledge, exploratory data analysis, and model results increases our confidence in our results.

The results in this report require the narratives about union violence over the past decade to be reevaluated. In particular, the story about trade union violence has mainly been a story told over time. Analysts usually emphasize increases and decreases in the reported number of killings. All parties to the debate tell a relatively consistent story about the most violent years, and it is generally accepted that the relative intensity of the violence has gone down since the early 2000s (as we saw in Figure 1). We believe that it is precisely because the stories have been largely consistent that a more subtle story has been missed. As the dominant narrative since 2004 has been about improvement (i.e., reductions in violence), the voice of some union groups, such as the ENS and the CCJ, about a relatively small number of missed killings has sounded weak, and almost stubborn. The results of this report confirm their important point – the real issue to clarify is the true magnitude of trade unionists killed in Colombia. These organizations who are closest to the victims are right to point out that some of the violence has been overlooked. What was not known, until now, is that the quantity of trade unionists killed and not recorded is substantial.

How might this require us to rework the existing narratives? For example, as mentioned above, in 2001 we estimate that the total number of homicides of union members was between 308 and 620 as compared to the 258 observed. The year 2003 may have been more deadly than 2004, although that was not previously apparent from observed data. The magnitude of killings was higher in earlier years than was previously thought. Interpretation of the impact of these higher numbers of killings is left to experts, but we wonder if perhaps these higher levels of violence in earlier years slowed union activity and affiliation in later years? Could it be that violence in later years is less because more unionists were actually killed earlier?

The more general lesson here is that underregistration is not homogeneously or randomly distributed. It must be estimated before data is used to make claims about patterns of violence. Using partial and inevitably biased convenience sample data will lead to erroneous results and interpretations. The changing rate of underregistration confirms that the raw data should not be used to make causal arguments. Only with estimates correcting for underregistration do researchers have the right input for causal models. A central example in this area is the question of whether the killing of union members is correlated in time and space with union activity.

5 Conclusion

When presented with a quantitative claim or pattern about violence, politicians, academics, activists, and ordinary citizens concerned with human rights violations should ask: are all the victims reported? In nearly all cases, the answer is probably not, as there are probably cases omitted from the records. How might the narrative that we understand to be true change if we knew the unknown cases? Any study that attempts to make quantitative comparisons that argue that violence is increasing or decreasing must take underregistration into account before quantitative analysis can be valid.

Underregistration has an impact on practically all of the conclusions previously reached using convenience samples. While the observed pattern about killings over time is largely the same, the magnitude of violence may have been worse in certain years than we observed. In fact, the years with the highest reported killings correlate with the years with the highest underregistration rate – making more killings linked to more uncounted deaths.

Although the patterns of estimated total killings of union members are similar to the patterns of the observed deaths, some subtle changes in the patterns may require adjustments to the interpretations. For one, between one-fifth and one-third of the total union-related homicides in 1999-2003 were unknown. How does knowing that those years were more violent than previously thought change the understanding of anti-union violence in Colombia more broadly? Santander also appears more violent, and the range of uncertainty for killings in this department indicates that we cannot rule out that Santander was as violent as Antioquia during the period studied here. Arauca is where the highest proportion of unionized workers were killed, with Putumayo and Caquetá not far behind. How does this fit into the broader narrative about violence in Colombia?

This study should be understood more as a starting point than as a conclusive report. It can serve as an input for future analyses, such as those arguing about proportions of responsibility, cause, and correlations with union activities. It can guide groups to know where extra data collection effort are important (for example, in Santander).

We have presented the existing data and estimated what it does not show us. We urge both technical and non-technical readers to study the methods appendix to understand how these numbers are calculated. We believe that there should be a debate about the appropriate statistical methods to correct underregistration and other biases in the databases documenting murdered union members.

Lastly, we hope to generate a broader reflection in the human rights community. If union killings are one of the most monitored violations in Colombia, and yet there are still in some areas and periods as many as 30% of the victims that have not been documented, what might this suggest for other types of violations? Any researcher attempting to quantify human rights violations, violence, or the effects of conflict must think carefully about underregistration and the potential for biased results and conclusions.

Appendix A: Methodology

Throughout this report we have discussed how under-registration in single datasets may cause errors in our conclusions about patterns of violence. Under certain special circumstances, access to two datasets (or systems) can provide a more reliable and less biased result than a single dataset [Sekar and Deming, 1949]. However, since two system estimates still rely on four strong assumptions, two systems are generally insufficient to correct the biases and unreliability of single datasets. Nevertheless, we introduce the mathematics behind multiple systems estimation (MSE) using the classic two-system estimate, outline the four assumptions required to calculate this estimate, then generalize to the three system case and explore the applicability and interpretation of these assumptions in the analysis presented in this report.

Two-System Estimates (Classic Capture-Recapture)

If every individual in a population of size N is equally likely to be sampled in system A, an unbiased estimate of the probability of "capture" by system A for a single individual is (sample size of A)/ (population size N). If another sample, B, is also taken from the same population, an unbiased estimate of the probability of capture by system B is (sample size of B)/(population size N). Furthermore, if the probability of being sampled in A is independent of the probability of being sampled in B (meaning that being sampled in A makes an individual no more or less likely to be sampled in B, and vice versa), then an estimate of the probability of being sampled in both A and B (call that group $M = A \cap B$) is ((sample size of A)/(population size N)) $\times$ ((sample size of B)/(population size N)). But note that, similar to the estimates of the probability of capture by each individual system, another estimate of the probability of being in group M is (sample size of M)/(population size N).

Numerically, that is

$$\widehat{Pr}(A) = \frac{|A|}{N} \tag{1}$$

$$\widehat{Pr}(B) = \frac{|B|}{N} \tag{2}$$

$$\widehat{Pr}(M) = \frac{|M|}{N} \tag{3}$$

where $|A|$, $|B|$, and $|M|$ represent the sample sizes of A, B, and M. Because of the independence assumption,

$$\frac{|M|}{N} = \widehat{Pr}(M) = \widehat{Pr}(A \cap B) = \widehat{Pr}(A)\widehat{Pr}(B) = \frac{|A||B|}{N^2} \tag{4}$$

Treating the true population size, N, as unknown, if we know A, B and M, then we can derive an estimate, $\widehat{N}$, of the unknown population size.

$$\frac{|M|}{N} = \frac{|A||B|}{N^2} \tag{5}$$

$$\widehat{N} = \frac{|A||B|}{|M|}. \tag{6}$$

Equation (6) is the two-system estimator for the unknown population size N.

Hidden within the above math are four underlying assumptions:

1. The samples come from a *closed system*: N must refer to the same population in each dataset.

2. The units are *homogeneous*: each individual in the population has equal probability of capture within any one system.

3. The systems are *independent*: the probability of capture by system A does not influence the probability of capture by system B (and vice versa).

4. *Perfect matching*: we are able to accurately partition all of the captured individuals into those only captured in system A, only system B, and those in both system A and B; all records referring to the same unit must be recognizable as such.

Analysts may determine that a single dataset is incomplete using dual systems estimates (e.g., Ball et al., 2007). But, with only two systems, there is no scientifically defensible way to correct the data following such a finding. Two systems are insufficient to determine the extent of the bias or to discover which of the two datasets is "more biased" (by whatever measure).

The first assumption, that the object of measurement– whether that is a population of persons in a country or a population of violent events that occurred in a state– is a closed system, is typically considered to be satisfied by homicide data. This assumption is generally unproblematic for data on violent events, because events that occurred cannot "un-occur" later. Although there is the possibility of 'false positives/negatives' (e.g., events that are later discovered to have occurred outside the time period of interest) we consider the risk and frequency of these to be low. This assumption can also be considered a coverage problem - as with any other sampling method, valid inferences can only be made about the target population, which is composed of members with the potential to be included on a list (for more in depth discussion of the implications of this assumption see [Manrique-Vallier et al., 2011, forthcoming] and Appendix B below).

The second assumption, homogeneity of capture probability, is unlikely to hold for any type of violence data. For example, persons with fewer social connections may be both more likely to experience a violent act, such as a kidnapping, and less likely to be recorded as victims; rural locations are more difficult to access than urban ones. Constructing two-sample estimates without accounting for different probabilities of capture leads to conclusions that may be biased.

The third assumption, independence of systems, is similarly difficult to meet. As with differences in capture probability, dependences between systems are impossible to account for in the two-system setting. A common example here is the difference between governmental and non-governmental organizations. Because different populations may have different levels of trust in the two organizations, reporting to one type of organization may imply that the witness is very unlikely to report to the other: the probability of capture in one system affects the probability of capture in the other. When two datasets are negatively correlated (like our example of a governmental and a non-governmental organization), two-system estimates will be inflated. If two datasets are positively correlated (perhaps two different government datasets), the estimates will be deflated.

The fourth assumption, perfect matching between systems, is the most computationally intensive part of the multiple systems process and involves matching records as accurately as possible using some unique identifier(s). See Appendix D for details on the matching process for data used in this report.

In the following Section we describe a model for estimating uncounted cases that does *not* rely on assumptions two and three. In more technical terms, the model is *robust to* violations of these assumptions and should therefore provide a much stronger estimate.

Estimates with three or more systems

Several researchers have developed techniques to correct for unequal probability of capture (violation of assumption two) and dataset dependences (violation of assumption three). These corrections are useful when three or more samples (datasets) are available [Sekar and Deming, 1949, Bishop et al., 1975, Darroch et al., 1993, Fienberg et al., 1999].

In order to account for unequal probability of capture, we use stratification, the division of the data into small sections that are more likely to have uniform probabilities of capture. It makes sense intuitively to stratify over both space and time, since both different geographic areas and different periods are likely to have different probabilities of capture. For a more theoretical justification for stratification, see Sekar and Deming [1949]. The estimates presented in this report are based on calculations stratified by time (year), location (department), and union sector.

Effective stratification requires that in each stratum, there be sufficient data in all systems, and sufficient overlap among systems. For example, we have found that in performing estimation with three systems, useful estimates are very difficult to achieve if there are no cases captured by all three systems (that is, the estimation fails if $Y_{111} = 0$; see explanation of cell counts below). As mentioned in Section 3.2, the structure of our data prevented stratification on more than one covariate at a time. Note that the observed values listed in Tables 2 and 3 are quite low for some departments and sectors, preventing further stratification along another dimension.

The third assumption described above requires that the fact of capture in one dataset does not affect the probability of capture in the other, i.e., that the datasets are independent. Several models that parameterize (i.e., that explicitly account for) non-independence of datasets have been suggested [Darroch et al., 1993, Agresti, 1994, Fienberg et al., 1999, Zwane and van der Heijden, 2007].

The application of this assumption becomes somewhat more subtle in the three-or-more system case. Depending upon which method is used to adjust for list dependency, typically strict independence is not assumed. Instead assumptions are made about higher-order independence and/or conditional independence. Additionally, though typically introduced as two separate assumptions (as we have here), heterogeneous capture probabilities and list dependencies can be treated as the same underlying problem. In other words, previously unidentified heterogeneity can often be detected through the appearance of list dependence, and adequately adjusted for through methods addressing dependencies. This will be elaborated following the introduction of log-linear models below.

A common solution to these paired challenges is to account for unequal probability of capture, to the extent possible, using stratification, and then to model residual dataset dependences using the log-linear model formalized by Bishop et al. [1975]. For three datasets, the basic problem is estimation of the missing cell in a $2 \times 2 \times 2$ table where each cell value Y describes the number of observations captured by a unique combination of the three datasets. $Y_{010} = n$, for example, means that n observations were counted in the second dataset only. Similarly, the cell value Y_{111} refers to the number of observations listed by all three datasets. For three datasets, eight log-linear models are possible. Where m_{ijk} is the expected cell count, u_1 is the parameter for dataset one, u_2 is the parameter for dataset two, and u_3 is the parameter for dataset three, one model suggests independence of the datasets:

$$\log(m_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} \tag{7}$$

Three models account for dependence between one pair of samples; they are analogous to

$$\log(m_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} \tag{8}$$

where, in this example, u_{12} is the parameter estimating the relationship between datasets one and two. Three further models account for dependence between two pairs of samples; they are analogous to

$$\log(m_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{23(jk)} \tag{9}$$

One model accounts for dependence between all three pairs of samples:

$$\log(m_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{23(jk)} + u_{13(ik)} \tag{10}$$

It should be noted that there is a different assumption implicit in these models - that the three-way interaction term (accounting for dependence between all three samples at once) is zero. This is a necessary limitation of the data - for any number l of lists, we must assume that the l-way interaction term in loglinear models is zero. This is the higher-order independence assumption mentioned above.

Several rules of thumb have been suggested for choosing the most appropriate model. The Bayesian Information Coefficient (BIC) balances goodness-of-fit and parsimony (i.e., models with fewer terms). The BIC is a logarithmic transformation of the chi-square: degrees of freedom ratio that better accounts for the "decreasing marginal returns" to degrees of freedom [Raftery, 1995, Hoeting et al., 1999]. For example, increasing from two to three degrees of freedom makes a great deal of difference to the quality of the model, whereas increasing from 202 to 203 degrees of freedom makes essentially no difference at all. Lower (i.e., more negative) BIC scores indicate models with the most appropriate ratio of goodness-of-fit to degrees of freedom, while BIC = 0 means that the model makes no improvement on the fully saturated model. It is worth noting here that many common statistical software packages calculate both the BIC and AIC (Akaike Information Criterion - another goodness-of-fit statistic). In particular, ecology methods tend to favor using the AIC since frequently in this field models contain many covariates and a smaller penalty for an increased number of model parameters is preferable [Burnham and Anderson, 2002]. For our purposes, the BIC is the preferred model fit statistic [Kass and Raftery, 1995].

Once a model has been chosen, $\hat{\mu}$, the maximum likelihood estimate of the intercept is exponentiated ($e^{\hat{\mu}}$) to estimate the total number of undocumented events (Y_{000}), which is then added to the number of observed events (n_{obs}) to estimate the total number of events ($\hat{N}$).

A Closer Examination of List Dependence and Capture Heterogeneity in the Context of the Results Presented in this Report

Analyses for this report were carried out using the R package `rcapture` [Baillargeon and Rivest, 2007]. A combination of loglinear models and stratification was used to adjust for potential dependencies between lists and capture heterogeneity. This section will discuss the applicability and detectability of potential violations of assumptions two and three described above in the specific analyses presented in this report.

As mentioned in the previous section, three-system estimates make it possible to employ MSE methods that are robust to some violations of assumptions - in particular, the combination of loglinear modeling and stratification used in this analysis allows us to examine potential capture heterogeneity and list dependence. It also modifies this latter assumption slightly - rather than assuming strict list independence we are assuming higher-order independence (i.e., that the three-way interaction term in our models is zero).

As mentioned in Section 1, analysts often explain differences in reported numbers of homicides by citing different definitions of what constitutes a 'case' for a given list. For example, one list may record all union member homicides whereas another list only includes union member homicides if that homicide is deemed to be a direct result of union activity. This may be considered an extreme example of a violation of assumption two - homogeneous capture probabilities[11]. In this case, we can expand our population of interest to include all union member homicides, regardless of underlying cause, but the second list clearly exhibits heterogeneous capture probabilities - homicides considered to be related to union activity have a different probability of inclusion on this list than homicides considered unrelated to union activity.[12] This is particularly problematic if we do not have a covariate indicating the believed motivation behind the homicide on which to stratify estimates. Left unadjusted, this heterogeneity could potentially result in biased estimates because the two lists have different inclusion criteria, limiting the amount of possible overlap. Unadjusted MSE calculations would simply interpret this as a small amount of overlap, and generate larger than necessary

[11] This may also be considered a violation of assumption one - that all systems refer to the same underlying population size N. This interpretation of differing case definitions is explored in the following Appendix B: Sensitivity Analyses.

[12] Arguably, if data collectors were able to perfectly identify the motivation behind homicides then a subset of the target population would have zero capture probability on one of the lists. We do not believe this to be the case in this example - that due to uncertainty in 'type' of homicide, deaths that are believed to be unrelated to union activity have a small, but nonzero probability of inclusion, even on lists focusing on homicides directly related to union activity. However, in the extreme example where a subset of the population truly had a capture probability of zero this would result in the coverage problem mentioned above in assumption one. See Manrique-Vallier et al. [2011] for a more in depth discussion of this problem.

estimates. Fortunately, this heterogeneity becomes apparent in the form of a negative correlation between these two lists, which we can detect using the loglinear models which become possible with the addition of a third list (or system).

We can also compare our model results with qualitative information about the data collection mechanisms. For example, as mentioned in Section 4, it is known that ENS and VP collaborated on data collection, particularly between 2001 and 2003. Overlap rates in Figure 2 also imply very comparable datasets for the years 2000-2003. This contextual knowledge and descriptive results are reflected in the pattern of models selected according to BIC (Table 4): for the years 1999 and 2005-2006 the BIC indicates the best fitting models are those that treat the lists as independent. In contrast, for 2000-2003 the best fitting models are those that adjust for dependence between VP and ENS ('x12' in Table 4) and also (for 2000-2002) ENS and CCJ ('x23' in Table 4). This consistency between contextual knowledge about the data collection mechanisms, exploratory data analysis, and model selection increases our confidence in MSE calculations.

Table 4: Estimates by Year

Year	Obs.	Est.	CI 95%	Formula
1999	95	124	(108 , 152)	$x1 + x2 + x3$
2000	187	237	(206 , 316)	$x1 + x2 + x3 + x12 + x23$
2001	258	386	(308 , 620)	$x1 + x2 + x3 + x12 + x23$
2002	255	323	(283 , 418)	$x1 + x2 + x3 + x12 + x23$
2003	133	173	(143 , 288)	$x1 + x2 + x3 + x12$
2004	117	126	(119 , 140)	$x1 + x2 + x3$
2005	80	86	(81 , 96)	$x1 + x2 + x3$
2006	87	90	(87 , 96)	$x1 + x2 + x3$
2007	44	49	(44 , 66)	$x1 + x2 + x3 + x12$
2008	65	80	(69 , 105)	$x1 + x2 + x3 + x12$

We can also sum each of the estimates across strata - although this does not provide a rigorous estimate of the total number of homicides, it does allow us to compare the implications of each of our stratified analyses. Summing estimates across years (Table 1) results in a total of 1,674 (95% CI (1,503, 1,845)[13]), across departments (Table 2) is 1,596 (1,394, 1,798)[14] and across union sectors (Table 3) is 1,638 (1,488, 1,788). The relative similarity of these sums indicates that we are probably adjusting for capture heterogeneity and list dependency adequately. If one (or more) of our stratified analyses resulted in a markedly different sum, this could indicate an unmeasured or uncontrolled source of heterogeneity. Although this is still a possibility, the similarity of results across three different stratification schemes makes it less likely. Lastly, it is important to note that this similarity across stratification schemes does not imply identical marginal homogeneity for years, locations, and union sectors. This would be the case if we fit identical models for each strata, but since the 'best' model differs across strata, we are adjusting for capture heterogeneity and implied list dependence differently for different strata.

Appendix B: Sensitivity Analyses

As noted in the previous section, the 'closed population' assumption is generally considered to be met by homicide data since events that occur cannot 'un-occur' later. However, the technical definition of this assumption is that each

[13]It is important to note that the error used to calculate these confidence intervals differs slightly from the method used to calculate confidence intervals for individual strata. The R package rcapture calculates a multinomial profile likelihood confidence interval for individual strata [Baillargeon and Rivest, 2007]. In contrast, we took the square root of the sum of the squared error terms for each individual strata to generate the confidence intervals reported here.

[14]As noted in Section 3.2, the small sample size and structure of available data made it impossible to calculate estimates for some departments and resulted in large amounts of uncertainty for other estimates. Only those departments for which estimates were presented in Section 3.2 contributed to the uncertainty in this sum.

system refers to the same underlying population size N. Again, as noted above, this can be considered a coverage problem - valid inferences can only be made about the target population and each member of the target population must have the *potential* to be included on a list.

In many applications, and specifically in the measurement of union violence presented in this report, the target population is defined by the inclusion (and exclusion) criteria used by the groups collecting the data. In other words, each system must make the same decision whether or not to 'count' an observed event. Although all three data collection agencies (VP, ENS, and CCJ) described their criteria as including killings believed to be directly related to union activity, the actual implementation of this criteria is inevitably subjective and may not be consistent across organizations. This implies that the definition of the target population (homicides related to union activity) and therefore the underlying population parameter N, may be slightly different for the different organizations.

In particular, during a meeting with ENS and CCJ, ENS pointed out that they were aware of specific records for which this was the case. As mentioned in Section 2.2, we know that ENS and VP collaborated during the early 2000s. As a result of this collaboration, ENS learned of homicides which they included in their dataset and VP observed but deemed to be unrelated to union activity and therefore excluded from their dataset. Ideally, we would like to be able to precisely identify these records. However, in the absence of this level of information, we can conduct the sensitivity analysis described below, randomly selecting records to drop or move, to determine if our estimates, and conclusions based on those estimates, are sensitive to potential variations in the definition of the target population.

The sensitivity analyses described in this section consisted of systematically altering the number of records assigned to each system. Recall that Figure 8 displays the number of events recorded in only one of the three systems (the outermost part of each of the three circles), in two of the systems (the overlap between any two circles) and in all three systems (the center where all three circles overlap). If, for example, CCJ and VP differed slightly in which homicides they determined to be directly related to union activity, then it is possible that some events were observed by both groups but only recorded by one (because the other determined that the event did not meet their inclusion criteria). This would a) result in slightly different target populations and underlying N for CCJ and VP and b) imply that some of the events recorded by only one system should either be excluded from analysis (to make the smaller target population common to both groups) or moved from the single list category to the two-list overlap category (to make the larger target population common to both groups).

This approach is precisely what we carried out in our sensitivity analyses. The quantity and structure of data did not allow us to analyze all three datasets at once - dropping or moving records from two systems at a time resulted in models that did not converge properly. However, we were able to systematically examine four cases, each driven by the question: What if ENS and VP or CCJ and VP observe the same homicide, but VP determines that homicide is not directly related to union activity?

1. Target population defined by VP $\Rightarrow$ *drop* proportion of records that only appear in ENS dataset.

2. Target population defined by VP $\Rightarrow$ *drop* proportion of records that only appear in CCJ dataset.

3. Target population defined by ENS $\Rightarrow$ *move* proportion of records that only appear in ENS dataset such that they are listed in both ENS and VP.

4. Target population defined by CCJ $\Rightarrow$ *move* proportion of records that only appear in CCJ dataset such that they are listed in both CCJ and VP.

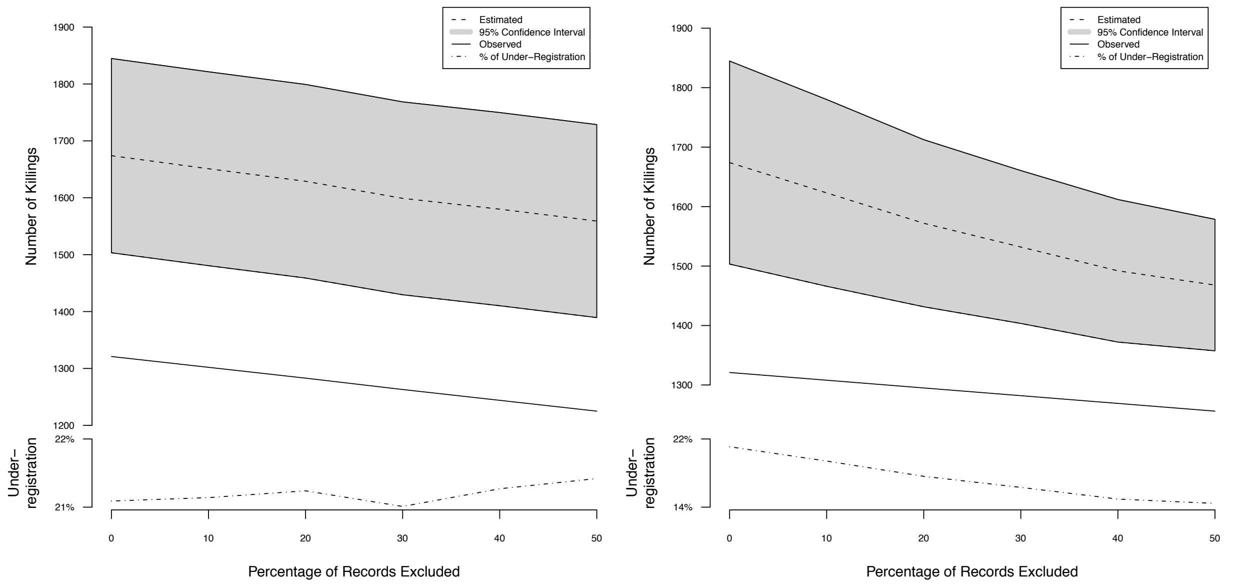
A similar set of comparisons could also be made for CCJ and ENS. However input from our partners at CCJ indicates that their inclusion criteria are much more similar to ENS's than VP's (i.e., the target populations as defined by CCJ and ENS are more likely to be similar). Therefore this third pair of potential comparisons was excluded from the sensitivity analyses since we believe the potentially more extreme differences between VP and ENS and VP and CCJ adequately describe any potential sensitivities in our calculated results.

For each scenario described above we re-calculated MSE estimates 5 times as we dropped or moved proportions of records in increments of 10% (e.g., dropping 10%, 20%, etc. of records that only appear in the ENS dataset). Each of these estimates was calculated by year, since the pattern of violence over time is primarily of interest. We then aggregated estimates for a total over time.

Figure 13 shows the estimated total number of homicides (dotted line), a nominal 95% confidence interval around this estimate (shaded region), the total number of observed records (solid line), and the estimated underregistration rate (dotted line at bottom of graph) as the proportion of records dropped from the ENS-only category (a) or CCJ-only category (b) is increased (scenarios 1 and 2 described above). Note that the upper limit of the x-axis (the maximum proportion of records dropped) is much larger than we consider plausible - we think it is extremely unlikely that half the records observed by the ENS or CCJ were also observed by the VP and not considered to be related to union activity. However, by including this extreme value we can examine the general behavior of our calculated estimates of the total number of union-related homicides. It is also important to note that in this scenario we are dropping records from analyses, resulting in a lower overall number of observed records (the solid line ranging from 1,321 to 1,225 (in Figure 13(a)) or 1,256 (in Figure 13(b))) as between 0% and 50% of either ENS- or CCJ-only records are dropped. As the total number of *observed* records decreases, the *estimated* total number of homicides naturally decreases as well, which can be seen in both (a) and (b) of Figure 13.

In general, dropping ENS-only records (Figure 13(a)) seems to have less of an effect on the estimated total than dropping CCJ-only records (Figure 13(b)). This is most noticeable by examining the dotted line indicating the underregistration rate - in Figure 13(a) this only varies between 21% and 22%. Meaning that even if some of the ENS-only records are 'miscategorized' (e.g., are records of homicides that are not related to union activity), we still calculate a stable, relatively high rate of *unobserved* homicides.

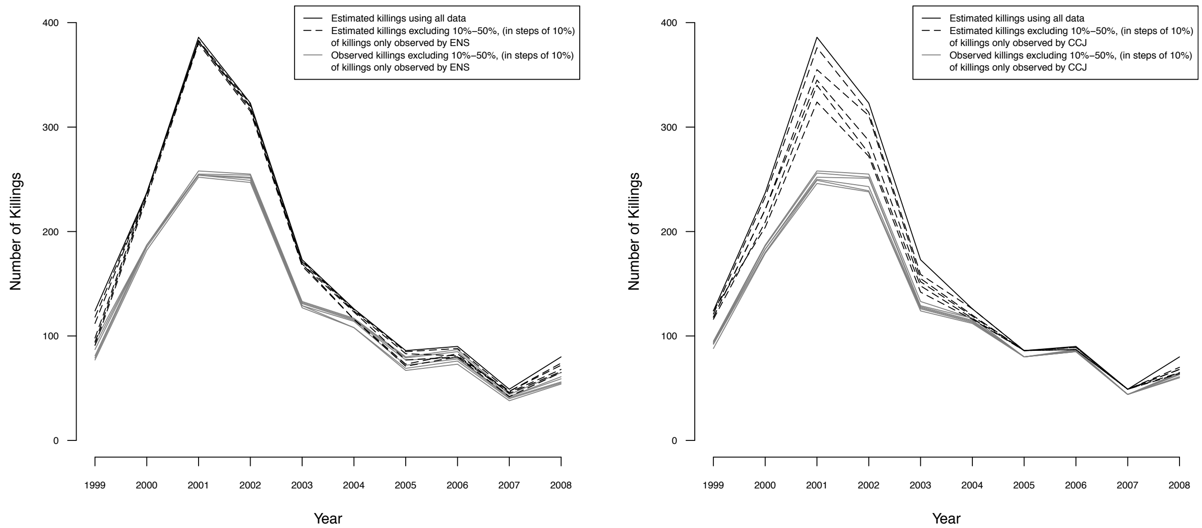
In contrast, when we start dropping CCJ-only records (Figure 13(b)) we see underregistration rates vary from just below 22% to just over 14%. Two things must be noted - 1) this reduced underregistration rate is only observed at levels of dropped records which we do not consider to be likely and 2) even this lower underregistration rate is statistically significant, again, meaning that even in the unlikely situation that a large number of CCJ-only records are 'miscategorized' we calculate a non-trivial rate of *unobserved* homicides.



(a) ENS (b) CCJ

Figure 13: Change in Estimated Total Number of Homicides and Underregistration Rate as Proportion of Records Omitted Increases

Figure 14 shows observed and estimated total homicides for each year using each of the datasets described above - dropping 10%, 20%, 30%, 40%, or 50% of ENS- (Figure 14(a)) or CCJ-only (Figure 14(b)) records. The grey solid lines in Figure 14 show the *observed* number of records for each year - remember as noted above that since in this scenario we are dropping records, the observed number decreases as we drop a larger proportion of records. The dashed black lines in Figure 14 indicate the *estimated* number of homicides. All of the dashed lines corresponding to estimated values in Figure 14(a) are clustered closely together, reinforcing the interpretation from Figure 13(a) that dropping ENS-only records, even a large proportion, has a small effect on the calculated estimates. In contrast, the dashed black lines indicating the estimates in Figure 14(b) are spread further apart, which coincides with the interpretation of Figure 13(b) that dropping CCJ-only records has a larger effect on calculated estimates. However, two things must be noted - 1) even in Figure 14(b) during years with the highest number of homicides, the estimates are well above the observed number of records, indicating that there are *unobserved* homicides during those years and 2) the overall pattern over time remains consistent, both within Figure 14 and Figure 7 (in Section 3.1).

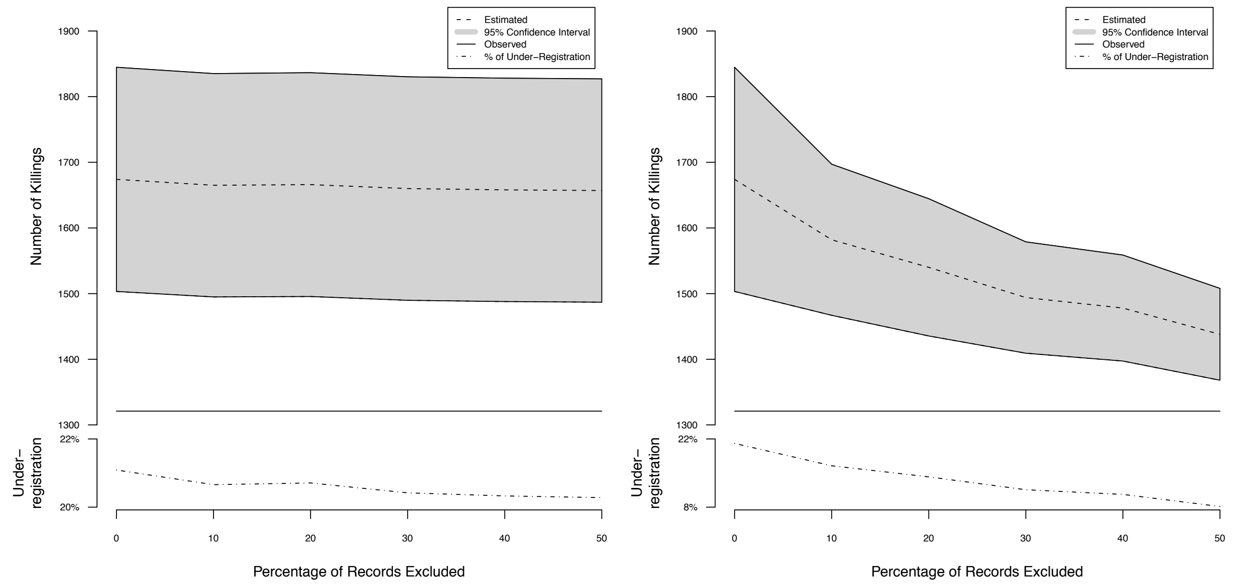


(a) ENS (b) CCJ

Figure 14: Change in Estimated Total Number of Homicides and Underregistration Rate as Proportion of Records Omitted Increases

Figure 15 shows the estimated total number of homicides (dotted line), a nominal 95% confidence interval around this estimate (shaded region), the total number of observed records (solid line), and the estimated underregistration rate (dotted line at bottom of graph) as the proportion of records *moved* from the ENS-only category (a) or CCJ-only category (b) is increased (scenarios 3 and 4 described above). Note that the solid line in Figure 15 is horizontal - in this scenario we are moving records from one category (ENS- or CCJ-only to ENS+VP overlap or CCJ+VP overlap) so the observed number of records remains constant (in contrast to Figure 13).

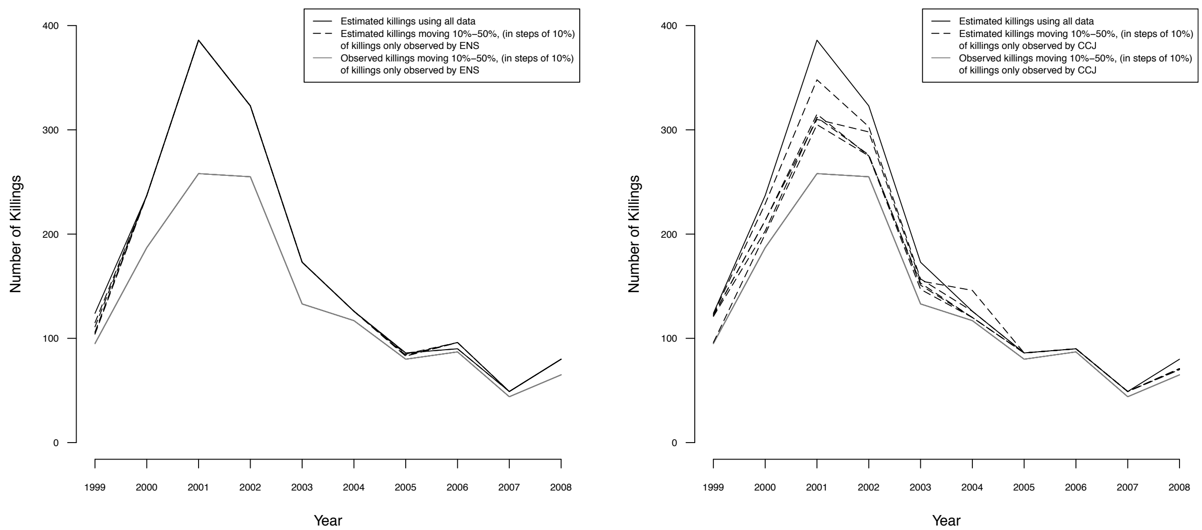
As in Figure 13(a), Figure 15(a) shows very stable estimates - even which a large number of ENS-only records are moved to the ENS+VP overlap category the estimated total number of homicides varies little. In particular, the underregistration rate only varies from 20% to 22%. Again, in contrast, Figure 15(b) shows more variability - this time the underregistration rate varies from 8% to 22%. It is worth noting that in the most extreme category where 50% of records are moved the estimated total number of homicides (1,438) is only slightly smaller than the estimated total number of homicides in the corresponding category where 50% of records are dropped (1,468 in Figure 13(b)). However, the underregistration rate is smaller (8%) when the records are moved versus when they are dropped (14%) since in the latter case the total number of observed records decreases too.



(a) ENS (b) CCJ

Figure 15: Change in Estimated Total Number of Homicides and Underregistration Rate as Proportion of Records Moved Increases

Lastly, Figure 16 shows observed and estimated total homicides for each year using each of the datasets described above - moving 10%, 20%, 30%, 40%, or 50% of ENS- (Figure 16(a)) or CCJ-only (Figure 16(b)) records to the ENS+VP or CCJ+VP overlap categories. As with Figure 14, this reinforces our interpretation of Figure 15 - the estimates in Figure 16(a) are so close together as to be nearly indistinguishable, whereas estimates in Figure 16(b) indicate the somewhat larger effect of moving CCJ-only records to the CCJ+VP overlap category. Again, we see the consistent pattern of violence over time, with peaks in homicides in the early 2000s and lower reported and estimated rates in more recent years.



(a) ENS

(b) CCJ

Figure 16: Change in Estimated Total Number of Homicides and Underregistration Rate as Proportion of Records Omitted Increases

In conclusion, we find our estimates of the total number of union-related homicides to be quite stable. Even in the highly unlikely scenario that a large proportion of records were 'misclassified' (e.g., that one or more of the organizations made different subjective choices regarding the inclusion of observed records) the estimated total number of homicides and especially the estimated underregistration rate remains remarkably consistent. Much like the consistency between our contextual knowledge, exploratory data analysis, and model selection, these results bolster our confidence that our analyses have adequately adjusted for the specific features of this data and, in this case, that our conclusions are not sensitive to plausible variations in the definition of the target population.

Appendix C: Data sources

Escuela Nacional Sindical (ENS)

The National Union School collects, classifies and analyzes information on violence and violations of the right to life, liberty and the physical integrity of Colombian trade unionists. They consider a human rights violation "any positive or negative behavior by which a direct or indirect agent of the state infringes, on anyone and at any time, one of the rights set out and recognized by the instruments that make up the international law of human rights"; and, secondly, those breaches of IHL by armed groups that violate the life, integrity and freedom of unionized workers who are part of the civilian population. Given the specificity of their work, they do not systematize all cases of breaches of IHL. In a country like Colombia, where levels of impunity surpass 90%, it is necessary to systematize and track particular cases of violence against trade unionists committed by unidentified perpetrators, both material and intellectual.

Observatory of the Presidential Program on Human Rights and International Law (VP)

The Observatory of the Presidential Program on Human Rights and International Law (in Spanish, *Observatorio del Programa Presidencial de Derechos Humanos y Derecho Internacional*) keeps a database on homicides of persons belonging to "vulnerable groups" throughout the country, such as mayors, former mayors, councilmen, journalists, Indigenous people, teachers, and union members. The information on these vulnerable groups comes from multiple sources, such as the Colombian Federation of Municipalities, The National Federation of Councils (Fenacon), the Foundation for Press Freedom (Flip), Indigenous organizations, the Ministry of Social Protection, the National Police, and the State security organization (Departamento Administrativo de Seguridad, DAS), among others. The Observatory also keeps a weekly log of press information that is the product of the daily review of national and regional newspapers, as well as of nationwide radio networks such as Caracol and RCN, downloaded from the Internet. The information gathered through these media relates to political activity, judicial activity, other pronouncements, acts of the State security forces and other State agents, actions of the illegal armed groups, and peace initiatives and demonstrations against violence.

Colombian Commission of Jurists (CCJ)

The research area of the Colombian Commission of Jurists (in Spanish, *Comisión Colombiana de Juristas*) keeps a database on sociopolitical violence with a daily and ongoing register on human rights violations and breaches of International Humanitarian Law. By "sociopolitical violence" the CCJ refers to acts that constitute attacks against life, personal integrity, and personal freedom caused by abuses of power of agents of the State; those politically motivated; those that derive from discrimination of politically marginal persons; or those generated by the internal armed conflict. At present, the information processed by the CCJ originates in the following sources: 20 national and regional daily newspapers; two weekly news magazines with national circulation; direct complaints gathered by the CCJ; denunciations addressed to other human rights organizations, both national and regional, that monitor the situation of human rights and humanitarian law; information from national authorities (Ombudsman's Office, Attorney General's Office, Prosecutor General's Office, and State security forces) and the publication Noche y Niebla of the Data Bank of the Center for Research and Popular Education CINEP /Justicia y Paz.

Appendix D: Matching

Matching is the identification of multiple records in a database that refer to the same event or individual. There are two types of matching done with records of victims for the purposes of statistical estimates; intra-system and inter-system matching.

One way to think about intra-system matching is de-duplication. This is done when the same act is reported to one dataset more than once. Duplicate records are fairly common; several family members may report a crime to one institution or the same act may be reported by multiple press sources. Inter-system matching is record linkage across different datasets. This is the case if one witness denounces a crime to two or more institutions – an NGO *and* a government institution, for example.

We carried out intra- and inter-system matching at the same time. The three datasets available for this study list the names, sex, date, location and trade union of victims. Some of the datasets also include information about alleged perpetrator. We used several sorting criteria to group together multiple records on the same victim into a unique "match group." Each match group contains one or multiple records that refer to that same victim. Records were considered a match if all of the following conditions were met:

- Names were identical or plausible typos or alternative spellings
- Dates of reported death were similar, preferably within a few days
- Departments were identical
- Union to which the victim belonged were similar or plausible abbreviations.

In uncertain cases, municipality, perpetrator, or consistency of reporting source may have been considered.

Matching was performed through two rounds of searching for ‘match groups’ - two or more records that met the criteria above to constitute an identical record. First, an analyst searched a single spreadsheet containing the pooled records from all three datasets for any possible matches. This spreadsheet was sorted many different ways, and for each sort the analyst assigned members of a ‘match group’ a common “match ID” number.

A second round of matching was performed to locate any additional matching records. First, all pairs of records were automatically identified in which

- the first *apellidos* of the two victims differed by an edit distance[15] of three or less;
- the first *nombres* of the two victims differed by an edit distance of two or less;
- the dates of death were fewer than seven days apart, accounting for possible swaps of the month and day; and
- the reported locations of the two deaths were in the same department.

We identified 2,064 pairs of records which met all of these conditions. We examined each of these record pairs, and marked the pairs as a match if we judged that they did indeed represent the same death. We thus matched 162 pairs of records not originally matched in the initial spreadsheet step.

Next, we reviewed all of the matches from both rounds to check for incorrectly-matched records. We computed a numerical score for each pair of matched records, according to the edit distance between each of the name components of the two records and the proximity of the reported dates of death. We then examined the pairs of records that scored the lowest on this metric, and corrected any mistaken matching decisions.

Finally, for each of these sets of records, we created a ‘representative record’ by combining the details of all the records in the set. The contents of this representative record were used for determining the date of death, economic sector, and other details needed for analysis. Usually, all of the records in a set agreed on these details, but there were occasional discrepancies between the reports. We resolved these discrepancies by choosing the value reported most often among the constituent records (the mode). When multiple records in the set came from the CCJ, the detail recorded by the CCJ only counted once in this comparison. When two conflicting details were reported an equal number of times, we chose the detail reported in the ENS record, or if no ENS record was present in the set, we chose randomly.

Acknowledgements

The authors thank Ronald Herrera, Jeff Klingner and Scott Weikart for their valuable contributions cleaning, processing, and matching the data for these analyses. We also thank Patrick Ball for editing each iteration of the report. We especially thank Shira Mitchell (PhD candidate Harvard University), Daniel Manrique (Postdoctoral associate Duke University), and Mauricio Sadinle (PhD candidate Carnegie Mellon University) for reviewing the report on a short timeline.

About the Authors

Daniel Guzmán, B.S., is a statistical consultant for the Benetech Human Rights Program. He has contributed to project design, partner training and data analysis in Colombia and Guatemala. He also conducted supporting data analysis for projects in Sierra Leone and Liberia. He received his B.S. in Statistics from the National University of Colombia.

Tamy Guberek, is a candidate for dual M.A. degrees in International and Global History from Columbia University and the London School of Economics. She served as the Latin America Field Coordinator for the Benetech Human

[15] The number of insertions, deletions, or substitutions required to turn one name into the other. For example, the names ‘Rob’ and ‘Bob’ have an edit distance of one.

Rights Program between 2004 and 2010.

Megan Price, Ph.D., is a statistician with the Benetech Human Rights Program. She has contributed statistical analyses to projects in Guatemala and Colombia. Dr. Price earned her Ph.D. in Biostatistics from the Rollins School of Public Health at Emory University.

About the Benetech Human Rights Program

HRDAG (the Human Rights Data Analysis Group) designs and builds information management solutions and conducts statistical analysis on behalf of human rights projects. With our partners, we make scientifically-defensible arguments based in rigorous evidence (http://www.benetech.org, http://www.hrdag.org).

This project was funded by core support to the Benetech Initiative from the the John D. and Catherine T. MacArthur Foundation, The Oak Foundation, and the Sigrid Rausing Trust, as well as project specific funding from Colombian Commission of Jurists and the European Union.

The Benetech Initiative
480 S. California Ave., Suite 201
Palo Alto, CA 94306-1609
tel: +1 650-475-5440
fax: +1 650-475-1066
Email: info@benetech.org
Web: http://www.benetech.org

References

Alan Agresti. Simple Capture-Recapture Models Permitting Unequal Catchability and Variable Sampling Effort. *Biometrics*, 50(2):494–500, 1994.

Peter Andreas and Kelly M. Greenhill, editors. *Sex, Drugs and Body Counts: The Politics of Numbers in Global Crime and Conflict*. Cornell University Press, 2010.

Sophie Baillargeon and Louis-Paul Rivest. Loglinear Models for Capture-Recapture Experiments. *Journal of Statistical Software*, 19(5), 2007.

Yvonne M. M. Bishop, Stephen E. Fienberg, and Paul H. Holland. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, 1975.

John Bohannon. The War in Afghanistan: Counting the Dead in Afghanistan. *Science*, 331(6022):1256–1260, March 2011.

K.P. Burnham and D.R. Anderson. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, 2002.

Ann Marie Clark and Kathryn Sikkink. Information Effects and Human Rights Data: Is the Good News about Increased Human Rights Information Bad News for Human Rights Measures? *Paper presented to the Institute for the Study of Human Rights, Columbia University, New York, January 26, 2011*, 2011.

John Darroch, Stephen Fienberg, Gary Glonek, and Brian Junker. A Three-Sample Multiple-Recapture Approach to Census Population Estimation with Heterogeneous Catchability. *Journal of the American Statistical Association*, 88(423):1137–1148, 1993.

Oeindril Dube and Suresh Naidu. Bases, Bullets and Ballots: The Effect of the U.S. Military Aid on Political Conflict in Colombia. *Center for Global Development*, 2010.

El Espectador. Cuentas a las Auc por Asesinar Sindicalistas. http://elespectador.com/impreso/temadeldia/articulo-253982-cuentas-auc-asesinar-sindicalistas, Feb 2011.

Escuela Nacional Sindical. No Cesan los Ataques Contra los Derechos Humanos de los Sindicalistas En 2010 Fueron Asesinados 51 Sindicalistas en Colombia, March 2011.

Stephen Fienberg, Matthew Johnson, and Brian Junker. Classical Multilevel and Bayesian Approaches to Population Size Estimation Using Multiple Lists. *Journal of the American Statistical Association*, 162(3):383–405, 1999.

Anita Gohdes. Different Convenience Samples, Different Stories: The Case of Sierra Leone. http://www.hrdag.org/resources/publications/Gohdes_Convenience%20Samples.pdf, 2010.

Tamy Guberek, Daniel Guzmán, Megan Price, Kristian Lum, and Patrick Ball. To Count the Uncounted: An Estimation of Lethal Violence in Casanare, 2010.

Madelyn Hicks and Michael Spagat. The Dirty War Index: A Public Health and Human Rights Tool for Examining and Monitoring Armed Conflict Outcomes. *PLoS Medicine*, 5(12):1658–1664, 2008.

Jennifer A. Hoeting, David Madigan, Adrian Raftery, and Chris Volinsky. Bayesian Model Averaging: A Tutorial. *Statistical Science*, 14(4):382–417, 1999.

R.E. Kass and A. E. Raftery. Bayes Factors and Model Uncertainty. *Journal of the American Statistical Association*, 90:773–795, 1995.

Bethany Lacina, Nils P. Gleditsch, and Bruce Russett. The Declining Risk of Death in Battle. *International Studies Quarterly*, 50(3):673–680, 2006.

Daniel Manrique-Vallier, Megan E. Price, and Anita Gohdes. Multiple Systems Estimation Techniques for Estimating Casualties in Armed Conflict. 2011.

Daniel Mejía and María José Uribe. Is Violence Against Union Members in Colombia Systematic and Targeted?, 2009.

Guillermo Correa Montoya. 2.515 o Esa Siniestra Facilidad para Olvidar: Veintiún Años de Asesinatos Sistemáticos y Selectivos Contra Sindicalistas en Colombia (1986-2006). *Escuela Nacional Sindical Área de Defensa de Derechos Humanos y Laborales*, July 2007.

Observatorio del Programa Presidencial de Derechos Humanos y DIH. Caracterización del Homicidio en Colombia, 1995-2006. 2009.

Megan Price and Daniel Guzmán. Comments to the article 'Is Violence Against Union Members in Colombia Systematic and Targeted?'. http://www.hrdag.org/resources/publications/Co-union-violence-paper-response.pdf, May 2010.

Adrian Raftery. Bayesian Model Selection in Social Research. *Sociological Methodology*, 25:251–266, 1995.

Leidy Sanjuán, Guillermo Correa Montoya, José Luciano Sanín Vásquez, Lina Paola Malagón Díaz, Pilar Peralta Díaz, and Harvey Rodríguez. Que os Duelan las Sangres Ignoradas: Impunity and Violation of the Human Rights of Trade Unionists in Colombia 2009-2010 and 2002-2010. *Cuadernos de Derechos Humanos*, 22, October 2010.

C. Chandra Sekar and W. Edwards Deming. On a Method of Estimating Birth and Death Rates and the Extent of Registration. *Journal of the American Statistical Association*, 44(245):101–115, 1949.

U.S. Labor Education in the Americas Project (USLEAP). Comments to the United States Trade Representative Concerning the Colombia Free Trade Agreement. September 2009.

Eugene Zwane and Peter van der Heijden. Analysing Capture-Recapture Data When Some Variables of Heterogeneous Catchability Are Not Collected or Asked in All Registrations. *Statistics in Medicine*, 26(1069-1089), 2007.