To Count the Uncounted: An Estimation of Lethal Violence in Casanare

Tamy Guberek, Daniel Guzmán, Megan Price, Kristian Lum, and Patrick Ball

Benetech Human Rights Program

Benetech
Technology Serving Humanity

February 10, 2010

1 Introduction

Casanare is a large, rural department in Colombia, with 19 municipalities and a population of almost 300,000 inhabitants located in the foothills of the Andes and on the eastern plains.[1] Multiple armed actors in the Colombian conflict have operated there, including paramilitary groups, guerillas and the Colombian military. Many people of Casanare have suffered violent deaths and disappearances.

But *how many* people have been killed or disappeared? For reasons of policy, accountability and historical clarification, this question deserves a valid answer. The Benetech Human Rights Program has used a statistical technique called Multiple Systems Estimation to estimate the total number of killings and disappearances in Casanare between 1998-2007. This report explains why it is often difficult to calculate an accurate accounting of the killed and missing, and why it is important to make sure these people are accounted for accurately. We then explain a methodology we have developed to estimate both the number of the known victims, and the number of victims who have never been counted.

Any accounting of lethal violence will be incorrect if we assume that any one dataset or combination of datasets contains a comprehensive count of killings and disappearances. Registries of violent acts kept by governmental and non-governmental institutions contain some, but not all, of the records of lethal violence. Organizations collecting this data may only have access to certain subsets of a population or geographic areas. Some reports of violent acts may be easier to locate than others and the resulting datasets will be biased toward those cases. How can we overcome these difficulties?

Correct answers about the number of killings and disappearances rely on statistical estimation to overcome the complex, incomplete patterns of reporting. Getting the numbers right is extremely important: appropriate estimates 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 and generates ongoing trauma for society. Victims who remain undocumented by any dataset become invisible, removed not only from their lives and the lives of their loved ones, but from historical memory.

Since we consider individual datasets to be incomplete, we prefer to use all available lists or datasets of killings and disappearances to generate statistical estimates. The estimation procedure used to calculate the magnitude of killings and disappearances for Casanare is called Multiple Systems Estimation (MSE).[2] MSE requires analysts to carefully review all known incidents in multiple lists, in order to determine whether some cases, either within one list or across lists, refer to the same victim. Matching cases that appear on more than one list allows statisticians to model the process by which violent acts are reported and to estimate the number of uncounted cases. MSE then uses the number of unique observations on each list in combination with the number of overlaps to estimate the total number of victims. Using a scientifically rigorous, transparent method to "count the uncounted" means that the results are less vulnerable to claims of partiality or bias.

The work presented in this paper builds on a previous study which estimated missing people in Casanare.[3] We chose to continue our research in Casanare for three reasons: 1) The line between killings and disappearances is often indistinct. Some people who are disappeared are presumed to be dead. In order to understand the magnitude of lethal violence affecting Casanare, we decided to analyze killings and disappearances side by side.[4] In all of the following analysis, we present results for killings and disappearances together so that readers can draw a comparison between

[1] DANE: Census 2005, Bogotá 2006.

[2] Chandra C. Sekar and Edwards W. 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.

[3] Guzmán, Daniel, Tamy Guberek, Amelia Hoover, Patrick Ball. 2007. "Missing People in Casanare."Benetech Human Rights Program. http://hrdag.org/resources/publications.shtml

[4] We have included killings and disappearances according to the most broadly used definition in the datasets provided to us. For deaths, this includes all homicides (the act of a human intentionally killing another human being) except suicide and accidents; for disappearances, this includes missing people where the fate of the victim is not known. The slightly different way each source defines these lethal acts is described by dataset in Section 3.1.

the pattern and magnitude of the two lethal acts. 2) Since the release of our 2007 report on missing people in Casanare, we have made important methodological improvements to our implementation of MSE. These advances allow us to include all of the available datasets and capture more precisely the range of uncertainty in the estimates. 3) We are integrating into this analysis new data shared with the Benetech Human Rights Program since the 2007 report. In this study, we have used information about victims of killings and disappearances provided by 15 datasets.

These 15 sources of data come from state agencies – including government, security, forensic and judicial bodies – and from civil society organizations. Using this data and our methodological developments, we estimate that there were between 3,944 and 9,983 killings in Casanare from 2000-2007. In the period from 1998-2005, we estimate that there were between 1,270 and 5,552 disappearances in Casanare. We present and discuss these estimates in more detail in Section 2. In Section 3, we describe the reported data and how it was processed for use in the analysis. We also show how descriptive summaries of individual datasets may be misleading. In Section 4, we draw some general conclusions. In Section 5, we outline areas where we plan to focus our future work. Lastly, we offer the methodological developments in technical detail in an appendix.

2 Results

In this section, we present estimates of the total number of killings and disappearances in Casanare between 1998-2007. Since one of our goals in this analysis is to calculate estimates by year, our period was determined by years for which there was sufficient data to determine these disaggregated estimates. By disaggregated, we mean estimates for the most specific time, geographic location and violence type that the data enable us to examine. To estimate deaths, we have enough data to calculate estimates for the time period 2000-2007 and for disappearances in the period from 1998-2005.

The results we get from Multiple Systems Estimation are a range of plausible values for the magnitude of total victims. Each specific value in this range has an associated probability that the value is correct given the data available. The number we present as the "point estimate" is the *expected value* of the range. The expected value is calculated as a weighted average of the range of plausible values; that is, each value is weighted by its associated probability.

When interpreting these estimates it is important to keep in mind that we are extrapolating from what is known and directly observed to what has not been observed and therefore remains unknown. We may accept the fact that some victims have not been included in the existing documentation, but without statistical methods, we have no way to determine how many are undocumented. Without a statistical technique to estimate a total beyond what is observed, we cannot evaluate arguments about the total magnitude of victims. Using MSE, we are narrowing the range of uncertainty from knowing nothing about the undocumented victims to a plausible range in which the total lies - a range in which we have a high degree of confidence. This range of uncertainty is expressed by the credible intervals surrounding the point estimate.

"Credible intervals" are based on a fully specified distribution, which means we are 95% sure that the correct number of violent acts is within the interval (given that the interval is based on our observed data). This is a more direct interpretation of uncertainty than given in traditional confidence intervals. Traditional "confidence intervals" tell us that if we conducted the study 100 times, 95 of these times we would expect the answer to fall in the reported range.

The 95% credible intervals reported in this analysis may appear to be large, but it is important to keep in mind that they capture the uncertainty associated with many datasets *and* with the use of multiple sets of statistical models required to calculate the estimate. In this project, we have adopted an open approach to the available data. We have deliberately avoided preferring some datasets over others, because choosing datasets is equivalent to making political choices about which are "best." Different combinations of datasets produce slightly dissimilar estimates. We have chosen to include all the combinations of datasets, and to examine the distribution of resulting estimates, in order to avoid making assumptions about which datasets might be best. The statistical consequence of our decision is that the uncertainty around the estimates (measured by the credible intervals) is substantial. We believe that these credible intervals provide an accurate reflection of the real uncertainty surrounding the estimate of the total number of lethal acts in Casanare.[5]

2.1 Stratification

Recognizing that Casanare is diverse in many ways, we stratified the department into smaller, internally homogenous sub-sections and sub-periods. We stratified by type of violence (death or disappearance), space (groups of municipalities) and time (year). There are two complementary reasons why we stratify. The first is to satisfy one of the methodological assumptions of MSE, that every case in the population has the same probability of being included in any one dataset (see appendix). Second, from a research perspective, a global estimate that encompasses the entire area of study may hide differences in the patterns and magnitude of killings and disappearances. Stratified estimates, on the other hand, allow us to have a more precise understanding of the differences among periods and regions.

Although we consider disappearances as well as killings to be lethal violence, we recognize that the method by which the two acts are perpetrated varies, and further, that they tend to be reported in differently. For this reason, we conduct the necessary calculations separately for the two kinds of violence.

For the spatial stratification that divides violent acts into the location where they took place, we did not have enough data to calculate a separate estimate for each municipality. We grouped the 19 municipalities of Casanare by similar conflict experiences based on guidance from local experts. We have also grouped the data based on neighboring municipalities.

The following municipalities were grouped together:

- Geographic Region *D - Center*: Yopal and Aguazul
- Geographic Region *E - Piedemonte*: Sacama, La Salina, Tamara, Recetor, Chameza and Nunchia
- Geographic Region *F - South*: Tauramena, Monterrey, Villanueva, Mani and Sabanalarga
- Geographic Region *G - Plains*: Hato Corozal, Paz de Ariporo, Pore, San Luis de Palenque, Trinidad and Orocue

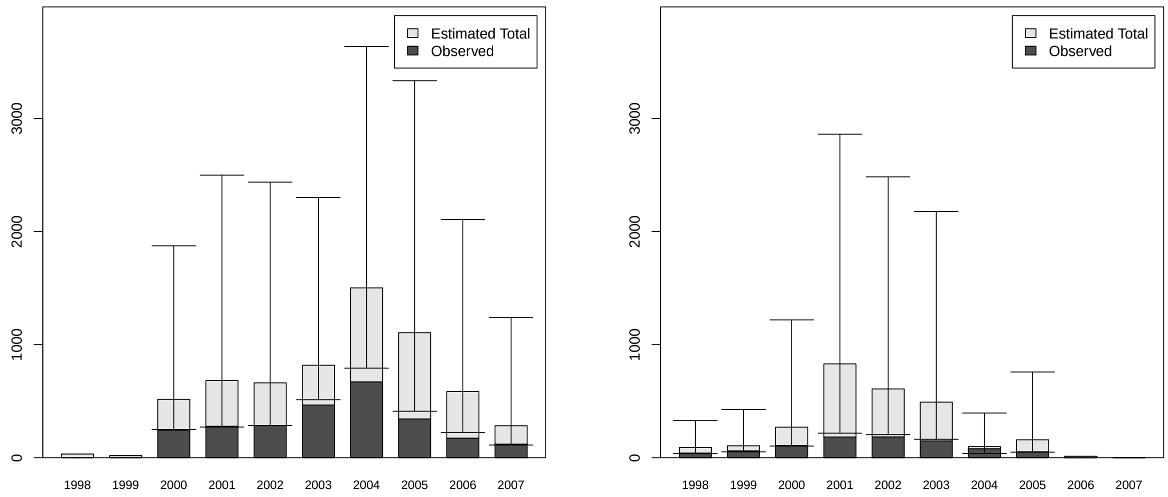
2.2 Estimates

2.2.1 Bar Graphs

First we present the results using bar graphs. The lower section of the bars in each of the following graphs represents the observed number of victims for that year. Stacked on top of this bar is the additional amount of under-registration such that the top of each bar indicates the total estimated number of victims. Under-registration refers to the number of unobserved killings or disappearances not recorded on any list. The lines (or "whiskers") indicate the 95% credible interval for each specific estimate.

Blank spaces in all graphs indicate years for which there was not enough data to generate an estimate. Blank spaces should not be interpreted as meaning there was no violence for that year and/or region. For more details on the meaning and implications of having not enough data, see Appendix: Methodology.

[5] It would be possible in a future project to make assumptions about which datasets are best, and to use only those datasets for estimation. The uncertainty around the estimates using a smaller set of datasets would be smaller than we have shown here. We want to be explicit that there is a tradeoff: making an additional assumption would reduce uncertainty. In this project, we have chosen not to make such an assumption.



(a) Killings (b) Disappearances

Figure 1: Observed and Estimated Killings and Disappearances, by Year

Figure 1(a) shows total killings in Casanare (observed and estimated) between 2000-2007. Based on the point estimate, there appears to have been a gradual increase in killings that peaked in 2004, then notably decreased. (Remember that the point estimate is the expected value from the range of plausible estimates.) However, we also note that the credible intervals are quite large and overlap across years. This means that although there is a discernible pattern among the point estimates we cannot conclude if such differences truly exist between years due to the inherent uncertainty in our estimates.

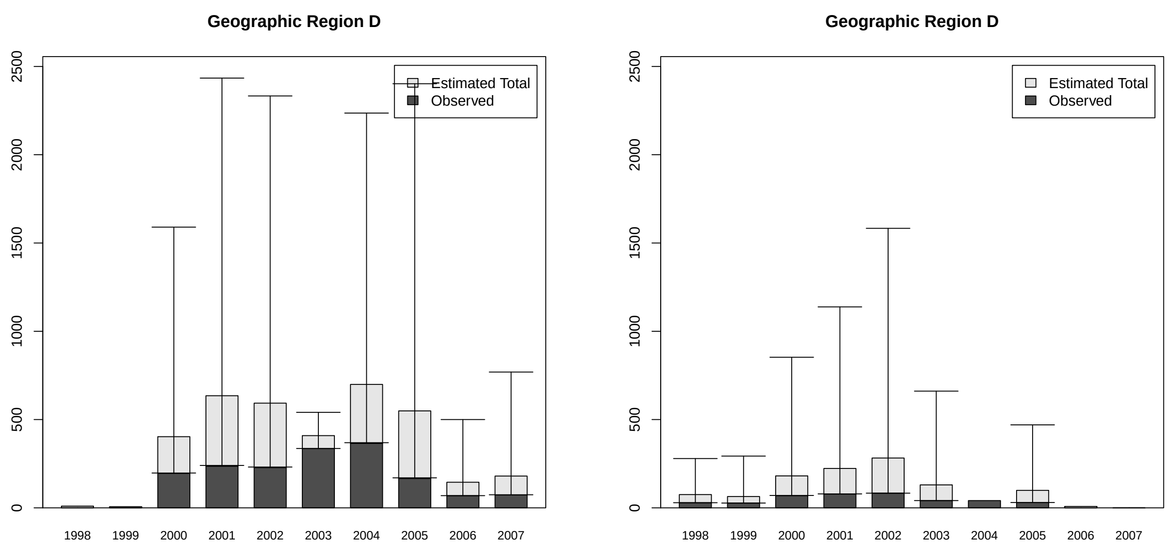
The pattern in the estimates of total killings follows the pattern of the reported data (lower part of bar). For example, the peak in the estimate for killings in 2004 is consistent with the story told by the underlying data, which also indicates a peak of killings in this year. What is interesting is that our estimate of under-registration (the number of unobserved victims) is also greatest in 2004. The method we use to analyze the reported data on victims is allowing us to estimate an even greater number of unknown victims who have never been counted. Of course, there is still a lot about these victims that we do not know.

The year 2007 is the one for which we have the least uncertainty, that is, the narrowest credible interval. The fact that the estimates for 2004 have more uncertainty could be the effect of changing documentation patterns. Although we have more data for 2004, the proportion of all deaths that were documented in 2004 may be less than the fraction of all deaths documented in 2007. It is also possible that the nature of deaths in 2007 was not the same as those in 2004, and that a higher fraction of 2007 deaths (relative to 2004) fit the criteria for inclusion by more of the datasets we used (see Section 3 for details regarding inclusion and exclusion criteria for each dataset).

In Figure 1(b), we see the number of observed and estimated disappearances by year. The observed values in 2001-2003 are similar, although the point estimates (the expected value from the range of plausible estimates) appear to be different: the point estimate is highest in 2001. However, the large amount of uncertainty, as measured by the 95% credible intervals, prevents us from concluding that the estimate in 2001 is significantly distinct from the estimates for 2002 and 2003. It is also worth noting that there is more uncertainty, i.e., larger credible intervals, for disappearances than for killings.

If we compare the estimates of total killings with the estimates of total disappearances, we observe contrasting patterns. Although disappearances and killings were approximately the same in 2001, total killings rose while disappearances likely fell.

In Figures 2–5, we present estimates disaggregated by killings/disappearances, year and geographic region. We have included the graphs for every region, although we only make observations about some of the graphs below.



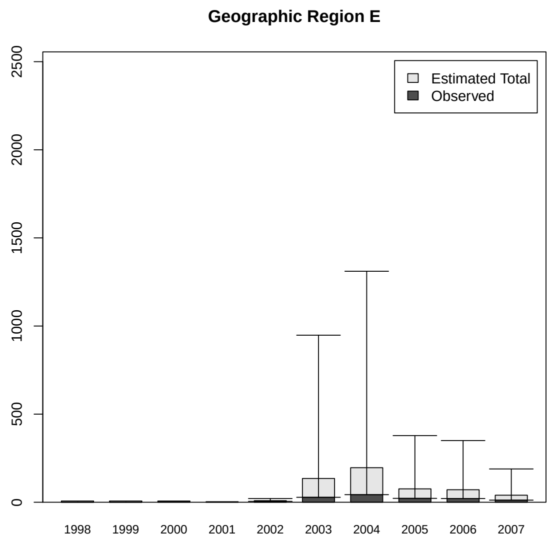
(a) Killings

(b) Disappearances

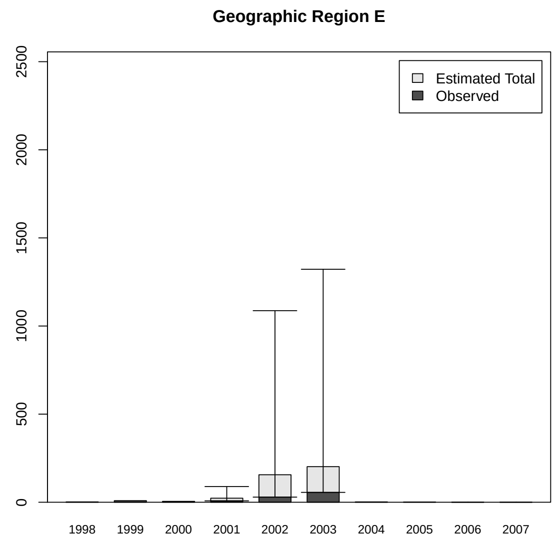
Figure 2: Observed and Estimated Killings and Disappearances, by Year for *Center*

In the region we have called *Center*, comprised of Aguazul and Yopal, we estimate high but persistent levels of violence. Other regions (presented below) have more variable patterns of violence.

In Figure 2(a), we estimate sustained high levels of killings, although with smaller estimates and credible intervals in 2003 and 2007. In Figure 2(b), we estimate a relatively sustained number of disappearances for *Center*, with numbers lower than those of killings in the same region.



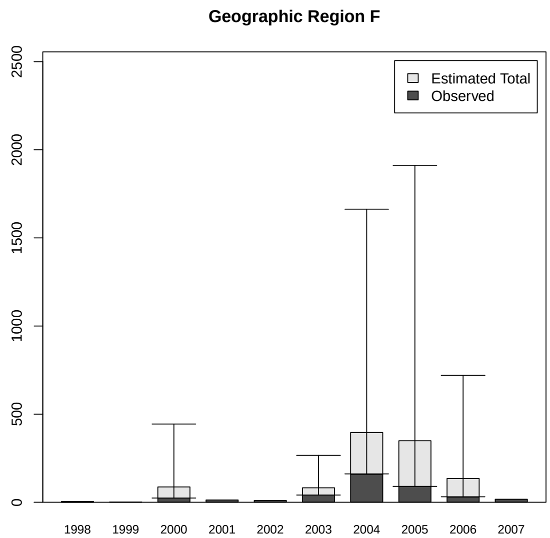
(a) Killings



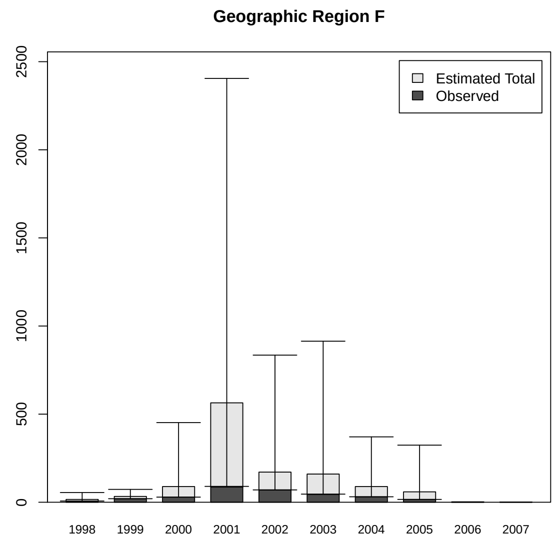
(b) Disappearances

Figure 3: Observed and Estimated Killings and Disappearances, by Year for *Piedemonte*

In Figures 3(a) and 3(b), we estimate relatively lower totals for the region we call *Piedemonte* than for other regions. However, there are two interesting observations: we can only make an estimate for both killings and disappearances in 2003; before that date, there were only sufficient reports to make estimates for disappearances and after 2003, there were only sufficient reports to estimate killings. An interesting pattern emerges when we calculate a per capita rate of violence by dividing the estimated numbers by the population size in each region below (see Figure 7(b)).



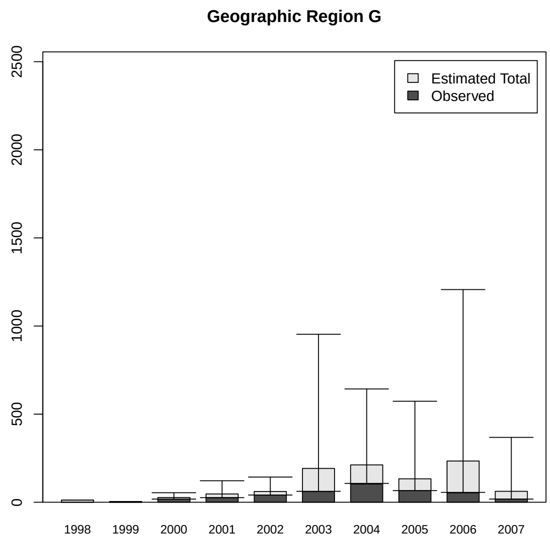
(a) Killings



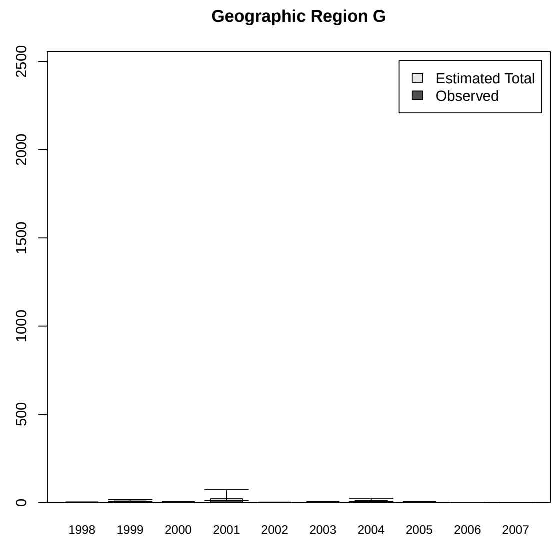
(b) Disappearances

Figure 4: Observed and Estimated Killings and Disappearances, by Year for *South*

Figures 4(a) and 4(b) show estimates for the *South* region of Casanare. We estimate a peak in disappearances in 2001. Comparing estimates of disappearances (Figure 4(b)) to estimates of killings (Figure 4(a)) in the same region, a trend emerges. In 2001-2002, there were few reports of killings, and therefore no estimate could be made. At the same time, reported and estimated disappearances were at their highest for the region – and for the entire department. It is clear that the patterns of disappearance are substantially different from the patterns of killings, region to region and year to year. In 2001, perpetrators may have been disappearing people instead of openly killing them; in later periods, the preference reversed, with more killings and fewer disappearances.



(a) Killings



(b) Disappearances

Figure 5: Observed and Estimated Killings and Disappearances, by Year for *Plains*

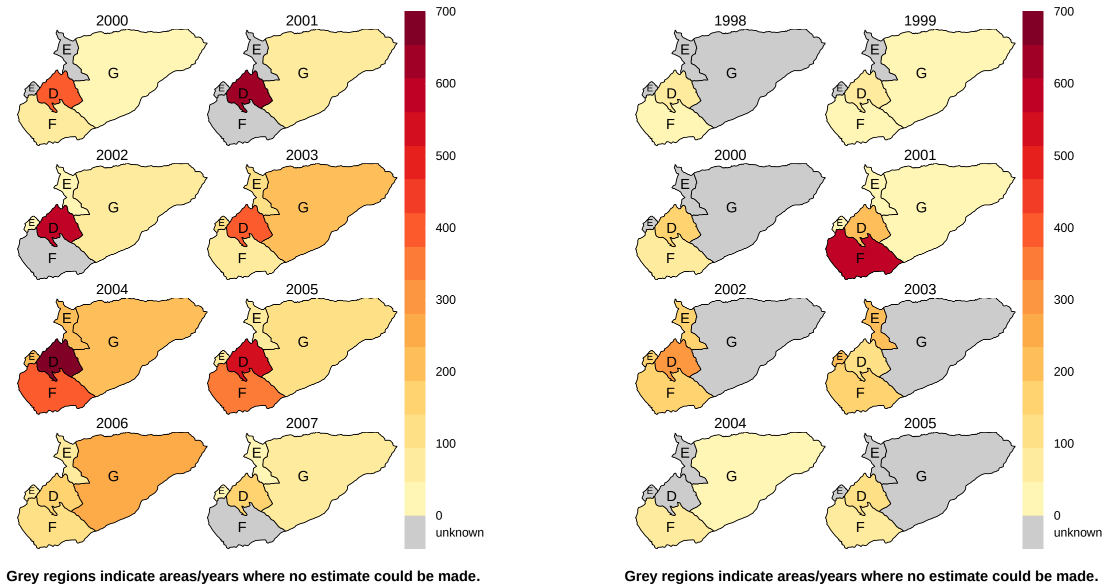
In the region we have called *Plains*, we estimate a steady low number of killings (Figure 5(a)). Few estimates could be made for disappearances (Figure 5(b)). Again, it is important to note that a lack of estimates does not necessarily mean a lack of violence. A lack of estimates means we know very little about the phenomenon of disappearances in this region.

These values are also listed in Table 1. These findings are discussed in more detail in Section 2.3.

2.2.2 Maps

The maps presented in this section use the same estimates presented in Figures 2(a) through 5(b), although we only present the point estimate (the expected value from the range of possible estimates) for ease of viewing. Complete point estimates and credible intervals can be found in Table 1.

The shadings indicate a range for the totals of killings or disappearances by geographic region within Casanare, allowing the reader to compare the intensity of the violence. Grey sections indicate that there is insufficient data available to generate estimates. Note that the years presented in Figure 6(a) for killings are 2000-2007 and in Figure 6(b) the years analyzed for disappearances are 1998-2005.



(a) Killings

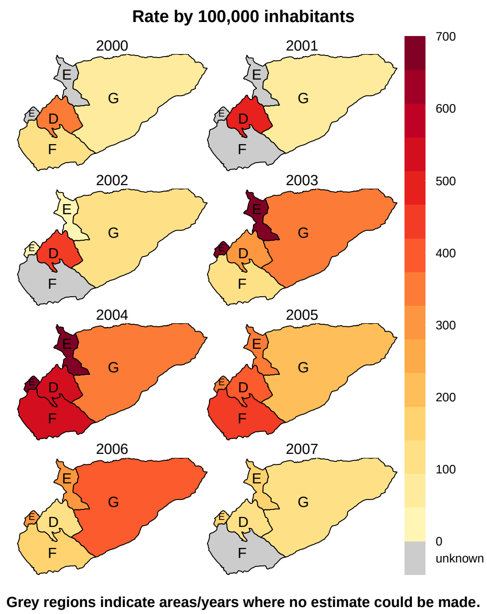
(b) Disappearances

Figure 6: Estimated Killings and Disappearances, by Region and Year in Casanare

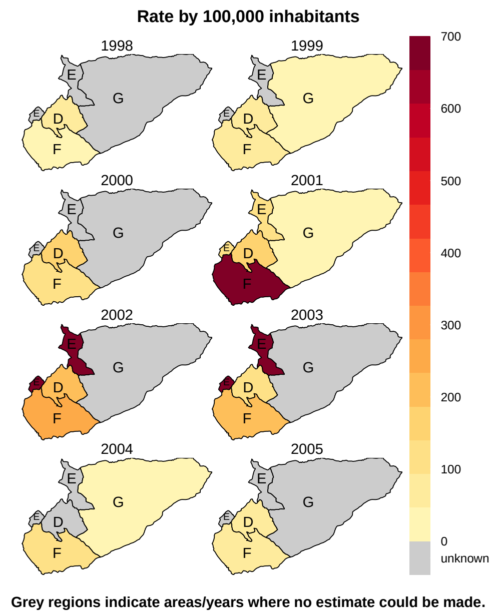
In Figure 6(a) we note that the highest numbers of killings were in geographic region *Center*. However, we also know that *Center* has the highest population density. Therefore, below, we present estimates as population rates, where we divide the point estimates of total victims by the total population by year.[6]

As we see in Figure 6(b), the most frequent place and period for disappearances was in the *South* of Casanare in 2001. The only year for which we could make estimates of disappearances for the entire department is 2001. For the *Piedemonte* and the *Plains* regions, we could only calculate estimates for 3 out of the 8 years.

[6] We used the projections for the total population of Casanare by year as calculated by the DANE. See (http://www.dane.gov.co/daneweb_V09/index.php?option=com_content&view=article&id=238&Itemid=121. Since we were unable to find the yearly projections disaggregated by municipality, we assigned the same proportions by municipality as the proportions in the 2005 census.



(a) Killings



(b) Disappearances

Figure 7: Population-adjusted Estimates of Killings and Disappearances, by Region and Year in Casanare

Figure 7(a) illuminates a pattern of killings in geographic regions E (*Piedemonte*) and G (*Plains*) which is not immediately obvious from Figure 6(a). Looking at all of Casanare for the whole period, geographic region E (*Piedemonte*) bore the highest population rate of killings in 2003 and 2004. In 2006, region G had highest population rate of killings for its population.

Figure 7(b) confirms the peak of disappearances in region F in 2001, as indicated in Figure 6(b). However, what is not obvious in the previous map is that the rate of disappearances in 2002 and 2003 in the *Piedemonte* region are among the highest.

2.2.3 Table

Table 1 below presents the data used in all the graphs and maps in the previous sections. For each year and region, it presents the counts of the unique observed victims, the estimated totals and the corresponding credible intervals, first for killings and then for disappearances.

Table 1: Killings and Disappearances

Geographic Region D - *Center*						
	Killings			Disappearances		
Year	Observed	Estimated	95% Credible Interval	Observed	Estimated	95% Credible Interval
1998	9	-	(NA , NA)	29	75	(29 , 279)
1999	6	11	(6 , 29)	27	64	(27 , 293)
2000	195	403	(197 , 1590)	69	181	(70 , 853)
2001	236	635	(240 , 2434)	78	223	(79 , 1138)
2002	228	593	(231 , 2333)	83	282	(83 , 1583)
2003	336	409	(336 , 541)	41	130	(41 , 661)
2004	365	699	(369 , 2236)	41	-	(NA , NA)
2005	166	549	(170 , 2402)	30	99	(30 , 470)
2006	68	145	(69 , 500)	8	-	(NA , NA)
2007	73	180	(74 , 769)	0	-	(NA , NA)
Total	**1,682**	**3,624**		**406**	**1,054**	
Geographic Region E - *Piedemonte*						
	Killings			Disappearances		
Year	Observed	Estimated	95% Credible Interval	Observed	Estimated	95% Credible Interval
1998	7	-	(NA , NA)	2	-	(NA , NA)
1999	7	23	(7 , 113)	9	-	(NA , NA)
2000	7	-	(NA , NA)	5	-	(NA , NA)
2001	3	-	(NA , NA)	8	23	(8 , 89)
2002	5	9	(5 , 21)	29	156	(29 , 1087)
2003	28	135	(28 , 948)	56	202	(56 , 1322)
2004	43	196	(43 , 1311)	2	-	(NA , NA)
2005	22	76	(22 , 378)	1	-	(NA , NA)
2006	21	71	(21 , 350)	1	-	(NA , NA)
2007	12	40	(12 , 189)	0	-	(NA , NA)
Total	**155**	**550**		**113**	**381**	
Geographic Region F - *South*						
	Killings			Disappearances		
Year	Observed	Estimated	95% Credible Interval	Observed	Estimated	95% Credible Interval
1998	4	-	(NA , NA)	7	16	(7 , 55)
1999	1	-	(NA , NA)	20	33	(20 , 73)
2000	24	87	(24 , 444)	29	89	(29 , 452)
2001	13	-	(NA , NA)	87	564	(90 , 2405)
2002	10	-	(NA , NA)	70	171	(70 , 835)
2003	41	82	(41 , 266)	46	160	(46 , 914)
2004	159	396	(161 , 1663)	31	89	(31 , 371)
2005	89	349	(90 , 1912)	16	59	(16 , 324)
2006	30	135	(31 , 720)	2	-	(NA , NA)
2007	17	-	(NA , NA)	1	-	(NA , NA)
Total	**388**	**1,049**		**309**	**1,181**	
Geographic Region G - *Plains*						
	Killings			Disappearances		
Year	Observed	Estimated	95% Credible Interval	Observed	Estimated	95% Credible Interval
1998	12	26	(12 , 88)	3	-	(NA , NA)
1999	4	-	(NA , NA)	5	8	(5 , 16)
2000	18	27	(18 , 54)	5	-	(NA , NA)
2001	26	47	(26 , 122)	10	21	(10 , 72)
2002	41	61	(41 , 143)	2	-	(NA , NA)
2003	61	192	(62 , 953)	6	-	(NA , NA)
2004	104	212	(107 , 643)	6	10	(6 , 24)
2005	66	133	(66 , 573)	6	-	(NA , NA)
2006	54	234	(56 , 1207)	1	-	(NA , NA)
2007	18	62	(18 , 368)	0	-	(NA , NA)
Total	**404**	**994**		**44**	**39**	

Remember that the lack of an estimate does not imply a lack of violence in a particular stratum. In the case of geographic region G (*Plains*), missing estimates cause an odd-looking total across years. Because we are only able to calculate an estimate for the number of disappearances in this region for three years, totaling these estimates results in a lower number than the total observed disappearances across this same time period. This is due to the fact that there were records of disappearances in 1998, 2000, 2002, 2003, 2005, and 2006 and yet in none of these years was data sufficient to calculate an estimate.

2.2.4 Comparison with 2007 report

It is useful to compare the estimates presented in this paper to those in our 2007 report. However, it is important to keep in mind that the estimates in the two studies were calculated using different datasets, different stratifications, and a different analytical method.

We are able to make estimates for different windows of time for the two studies. The only sections of our 2007 report that overlap with this study are the global estimates for all disappearances in all of Casanare during 2001–2004 and 1998–2005. In our previous study, the point estimate for all disappearances during the first period (2001–2004) was 821 within a 95% confidence interval of (713, 929). Our current methods estimate the number of disappearances across all of Casanare for this same period to be 2,028 within a 95% credible interval of (840, 4,700). On the one hand, these estimates are conceptually different, and each point estimate falls outside the confidence or credible interval from the other study. On the other hand, the credible interval calculated for this report (840, 4,700) is quite large and overlaps with the confidence interval calculated in the 2007 report.

For the longer time period, 1998–2005, our previous report estimated the number of disappearances across all of Casanare to be 1,239 within a 95% confidence interval of (1,088, 1,390). Our current report estimates the number of disappeared for this same period to be 2,653 within a 95% credible interval of (1,270, 5,552). Again, the point estimates appear to be quite dissimilar, but the intervals overlap.

We conclude that although we have collected additional data, there are more undocumented disappearances across Casanare than we originally estimated. As shown by the large credible interval accompanying this estimate, much remains unknown about disappearances during this time period. We believe that we underestimated the uncertainty of the estimate in the 2007 calculations, and the 95% confidence intervals are consequently too narrow. Despite the size of the credible intervals in our current report, we believe they reflect a more accurate accounting of the uncertainty in the data.

2.2.5 Total Lethal Acts

As noted in the previous section, the overall estimate of the total number of disappearances for all of Casanare between 1998 and 2005 is 2,653 within a 95% credible interval of (1,270, 5,552). The observed total number of disappearances for all of Casanare during this time period is 872, implying that between 398 and 4,680 disappearances are likely to have gone unreported to any of the 15 data sources used in this report.

Between 2000 and 2007 we estimate a total of 6,215 killings in all of Casanare, within a 95% credible interval of (3,944, 9,983). During this same time period 2,629 killings were reported, implying that between 1,297 and 7,336 killings are likely to have gone unreported to any of the 15 data sources used in this report.

In Section 3.1, we outline some of the challenges involved with data collection, which may explain why so many killings and disappearances go unreported.

2.3 Discussion

Our central conclusion is that a large under-registration of killings and disappearances in Casanare persists, even after substantial data collection efforts. There is still a great deal of uncertainty about the exact totals. In spite of the uncertainty, we have been able to outline certain patterns based on the time, space and type of violence. The peak in overall disappearances in 2001 largely took place in the *South*, precisely at the time when reported killings in the same region was at the lowest point. The results show that killings have been widespread in Casanare, reaching a peak in 2004. This large number of killings in 2004 was consistent across the regions of Casanare except the *Plains*, where we estimate higher killings in 2006. Furthermore, in regions where estimated totals appeared to be low (e.g., E and G), the population-adjusted maps reveal a relatively higher toll on the population. In particular, the *Piedemonte* region had some of the highest rates of disappearances in the department proportional to their population size.

Viewing patterns and magnitude of killings and disappearances side by side raises questions about the perpetrators' strategies and the dynamics of the conflict. Why would disappearances vary inversely with killings, as we saw particularly in the *Piedemonte* and the *South* regions? Why would we tend to see the most disappearances when and where we see the fewest killings, as in the *South* in 2001 and in *Piedemonte* in 2002? And vice-versa, why are there so few reported disappearances in *Center* when killings are probably at their highest? Why are there elevated killings in the *Plains* in 2006? Researchers from other disciplines may find it useful to combine the questions stemming from the quantitative patterns with their qualitative knowledge to gain a better understanding about lethal violence in Casanare over the past decade.

Estimating the number of killings and disappearances that were unobserved reveals the invisible dimensions of lethal violence in Casanare. As we saw in the estimates of totals, there are likely 70% more killings in Casanare than reported by all the lists combined in the period of 2000-2007. There are likely more than twice as many disappearances as those reported for the department, ranging from 1,270 - 5,552 in the period 1998-2005.

Going forward, we recommend that the greatest efforts should be channeled toward additional data gathering about lethal violence by all of the organizations in Casanare. A higher coverage rate will improve the ability of statisticians and social scientists to make more precise estimates, that is, to decrease the uncertainty around those estimates. Naturally, these data collection efforts will improve as families of victims and civil society in general develop more trust and security in the reporting process, more likely in the absence of conflict.

We emphasize that the rate of undocumented victims of lethal violence found in Casanare does not necessarily represent the rate that could be found in all of Colombia, if data were available. This analysis demonstrates that no single dataset of violent deaths accurately records the total number of persons likely to have been killed in Casanare. We explore this topic further in Section 3.

3 Data

3.1 Data Sources

Collecting data on violence is difficult. It involves security risks, trust-building, and logistical, financial and human resources. Many courageous and hardworking institutions engage in this work despite the challenges involved. It is natural that in a place as violent as Casanare that some groups will learn about some victims and other groups will learn about other victims. Some groups will learn about the same victims and some victims will be hidden from all data collection efforts. In order to better understand the data used in this study, we will briefly present the scope of the data collection effort of each organization according to its mandate used in this study.

All datasets collected by these groups were shared with us or with our partner EQUITAS for use in quantitative data analysis. Data was shared in its raw form; all the analysis and conclusions in this report were carried out exclusively by the Benetech Human Rights Program and do not reflect the opinions of the organizations that originally collected the data.

Table 2: Summary of Sources

Organization name	Violence Type	Period
National Institute of Forensic Medicine - Deaths	Killings	2000-2007
Attorney General's list of the Disappeared	Disappearances	1998-2005
Human Rights Observatory of the Vice Presidency	Killings	2004-2007
National Police (DIJIN)	Killings	2003-2006
CINEP	Killings and Disappearances	1998-2007
Fondelibertad	Disappearances	1998-2005
Colombian Commission of Jurists	Killings and Disappearances	1998-2007
Colombia-Europe-US Coordination	Killings	2005-2007
National Institute of Forensic Medicine - Disappeared	Disappearances	1998-2005
Attorney General of Santa Rosa - Disappeared	Disappearances	1998-2005
Families of Victims Organizations	Disappearances	1998-2003
Gaula	Disappearances	1998-2005
Technical Investigative Body of the Attorney General's Office	Disappearances	2003-2005
Equitas	Disappearances	2000-2005
País Libre	Disappearances	1998-2004

Each institution collects data according to their institutional mandate. Some groups, such as the Colombian Commission of Jurists (CCJ), monitor human rights violations. Some groups document victims for the purposes of humanitarian identification, such as Equitas. The *Instituto Nacional de Medicina Legal y Ciencias Forenses* (INMLCF) documents the cadavers they inspect which are victims of violent deaths. The Colombian National Police track homicides and the Prosecutor General (*Fiscalía*) records disappearances from denunciations and for judicial processes. (For more information about the sources, please see the sources appendix in the Spanish version of this document.) We should keep in mind for the purposes of this study that the victims included in each dataset vary according to these mandates.

The initial universe of the data we reviewed included a broader set of data classification, including all records in Casanare classified as either a disappearance, a simple kidnapping, an extortion-related kidnapping, a freed kidnapping, a death, a hostage or a detention. We excluded all suicides and accidents. Then we matched the data, as described in Section 3.2: that is, we checked each record from each database against all the other databases to determine whether the same person appeared in multiple databases. After matching, records classified as detentions, hostages and extorsive kidnappings were kept only when they were matched with another record classified as a death or disappearance. All other records classified as detentions, hostages and kidnappings were dropped.

In Table 3 the total records column indicates how many records of killings and disappearances were included in each dataset; the next column limits these records to those found in the period of study, 1998-2007; the final column, Uniques, indicates how many records in the period of study from each dataset were found only on that list, not in any other dataset. For example, there were 2,168 total records in the INMLCF deaths dataset, 2,085 that fell in the time period of 1998-2007, and 1,420 victims in that period that were not reported by any other list.

Table 3 makes clear that each individual dataset is incomplete. However, most datasets contribute new information: the majority of datasets in Table 3 contain at least some events that are documented only in that dataset. See Section 3.4 for more descriptive comparisons of the datasets.

Table 3: Total and Unique Reported Killings and Disappearances

Organization name	Total Records	1998-2007	Uniques
National Institute of Forensic Medicine Deaths	2168	2085	1420
Prosecutor General's list of the Disappeared	1313	659	387
Human Rights Observatory of the Vice Presidency	528	501	284
National Police - DIJIN	825	825	221
CINEP	338	274	97
Fondelibertad	332	312	67
Colombian Commission of Jurists	250	217	51
Colombia-Europe-US Coordination	77	72	30
National Institute of Forensic Medicine Disappearances	193	172	10
Prosecutor General of Santa Rosa	169	163	4
Families of Victims' Organizations	52	51	1
Gaula (Anti-extortion Unit)	128	111	1
Technical Investigative Body of the Prosecutor General's Office	36	36	0
Equitas	28	23	0
Pais Libre	9	9	0

3.2 Record Linkage

Record linkage - or "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* the Prosecutor General's Office, for example.

We carried out intra- and inter-system record linkage at the same time. Each dataset available for this study lists the names, sex, date, location and violence type. Some of the datasets also included information about alleged perpetrator, about physical characteristics of the victims and other variables. This means that each dataset included different variables. Prior to record linkage, we chose a minimum, common set of variables for all datasets. Next we combined all the datasets into one large dataset, keeping a unique identifier for each record by dataset.

The Benetech Human Rights Program 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. The records that we believed to refer to the same person sometimes matched exactly. Other times, they matched on at least the victim's first name, their first last name and their second last name. Dates and locations were either exact or varied by a few days. Sometimes month and day digits appeared to be inverted but all other information was the same.

In order to determine the sensitivity of the de-duplication and inter-system record matching, the data were matched by two human matchers. We then evaluated the similarity of their two different groupings of the records. The two matchers started with the same set of automatically identified matches, which they modified by identifying new matches and un-matching erroneously grouped records.

We quantified the outcome of this matching by counting the numbers of matched pairs of records which the matchers did and did not agree on. For example, let's say there were records of a Juan Garcia in five separate datasets. After looking at the details of the records' dates and locations, the first matcher grouped all five records together while the second only grouped four of them and said the fifth was a different disappearance. There are six *pairs* of records on which the two matchers agreed. We calculated the matchers' rate of agreement. This is the number of times they agreed on the correspondence of two reports of a lethal act divided by the number of times they could have agreed. Using this calculation, we found that the two matchers agreed 91.1% of the time. There were 280 pairs of records matched by the first matcher but not the second, 149 pairs matched by the second matcher but not the first, and 4,389 pairs of records matched by both of them.

This high rate of agreement between the matchers suggests that the outcome of the matching does not depend strongly on who does it. This reliability in matching also suggests that there is enough detail in the data to reliably connect separate reports of the same lethal act together and therefore determine the rates of overlap among the different data sources.

3.3 Contradictions in Match Groups

Sometimes, records that were clearly referring to the same victims still contained contradictory information. For example, a person may be categorized as "disappeared" in one dataset and "dead" in another dataset. Similarly, the date or place of the disappearance may vary due to imprecisely reported information or typographical errors in the data entry.

There are three specific areas where records can contradict each other: contradictions between violence type, municipality and year. We created the following hierarchy of rules to resolve contradictions within what we call a match group – groups of records from multiple datasets that we think refer to the same person:

Where records differed on violence type within a match group:

- If at least one record was classified as a death, the act represented by the group was considered a death
- If at least one record was classified as a disappearance and there was no death in the group of records, the act represented by the group was considered a disappearance
- Records classified as detentions, hostages and extorsive kidnappings were kept only when they were matched with another record classified as a death or disappearance. All other records classified as detentions, hostages and kidnappings not considered "lethal" and were therefore dropped from the study.

Where records differed on municipality within a match group:

- If one record had a missing value for municipality and the other had a valid municipality, we chose the valid municipality;
- If there were two or more municipalities, we chose the one most frequently reported;
- If one record had the capital, Yopal, for municipality and the other record had a different municipality, we chose the non-Yopal municipality;[7]
- If two records had different municipalities, and neither was Yopal, we chose the municipality at random.

Where records differed on year within a match group:

- If there were two or more years reported, we chose the one most frequently reported.
- Within the most common years, we chose the latest one.

These rules were applied prior to calculating estimates.

[7] The assumption here is that since it is likely that people denounce the event in the capital city Yopal, where most organizations have an office, it is probable that one organization recorded Yopal as the place of denunciation rather than the place of disappearance.

3.4 Descriptives

Descriptive statistics describe and summarize variables in a collection of data. They are useful for understanding observed data, but they cannot be used to make inferential conclusions about the entire population.

In this section, we present descriptive statistics using the combined data from the 15 datasets, after they were de-duplicated, matched and contradictions in multiple reports were resolved. We note the challenges of trying to understand killings and disappearances in Casanare relying solely on observed data.

3.4.1 Counts of Victims by Time and Space

The graphs below of act type by year are the same as the lower bar in the year estimates in Figures 8 and 9. We present them here to emphasize that this is the pattern of *reported* victims, and to distinguish reports from reality. As we see in Figures 8 and 9, without the estimates, we would have a flawed understanding not only of the magnitude of persons killed and disappeared, but we would have also missed the likely elevated effort of perpetrators to disappear victims in 2001.

Figure 8: Counts of Reported Disappearances by Year included in all Datasets

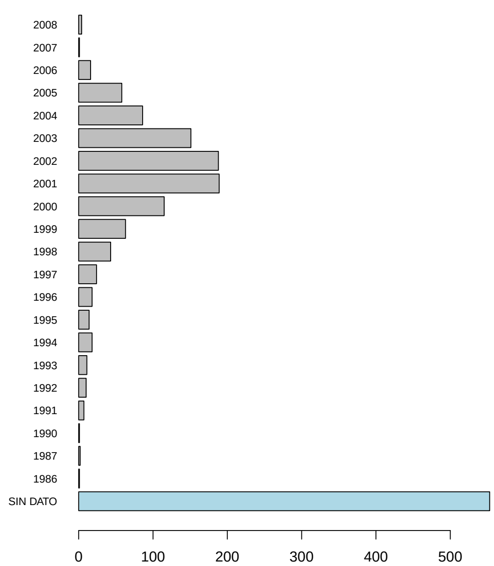


Figure 8 shows almost the same count of disappearances in 2001-2002, but as we saw above, we estimate higher unreported disappearances in 2001. We also note the high number of records with missing date information. The disappearances with missing date of disappearance must be excluded from the estimation because we cannot determine whether they occurred during the period of study.

Figure 9: Counts of Reported Killings by Year included in all Datasets

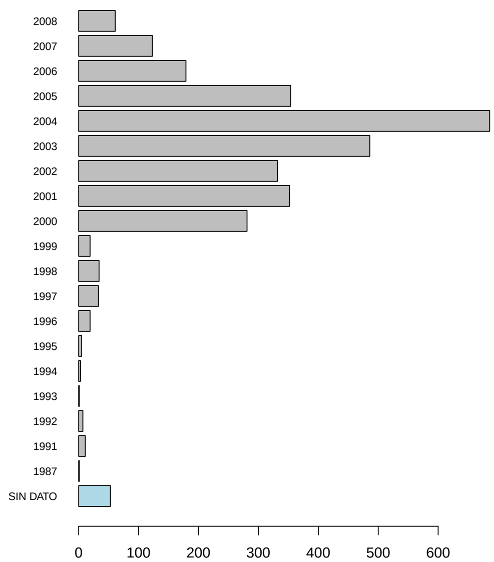


Figure 9 shows a similar pattern for killings as the estimated pattern. Note that unlike disappearances, relatively few killings are missing the date of the killing.

Next we present the counts of reported victims by municipality. It is useful to see the underlying data that was used to make the estimates by geographic region, especially since the estimates could not be calculated by municipality. We see in Figures 10 and 11 that Yopal has the highest counts for both killings and disappearances, but there are many records missing location information, especially for disappearances.

Figure 10: Counts of Reported Disappearances by Municipality included in all Datasets

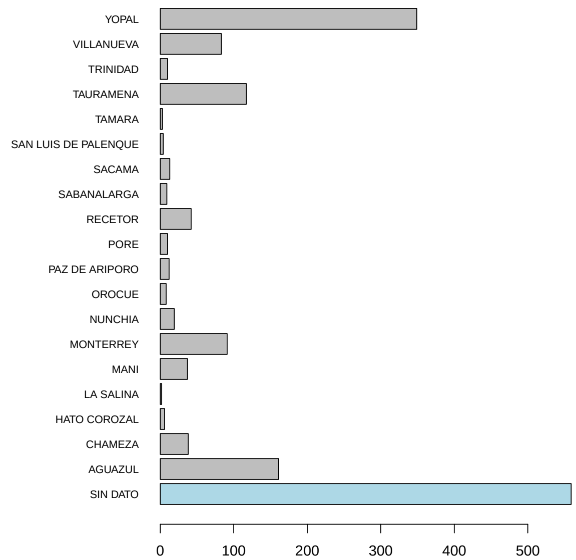
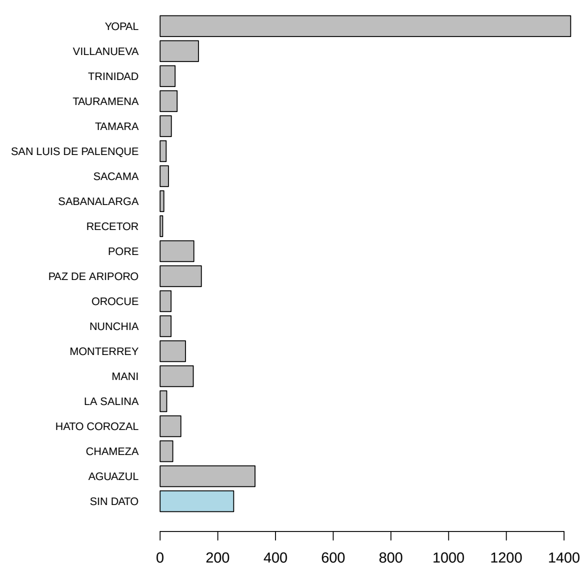


Figure 11: Counts of Reported Killings by Municipality included in all Datasets



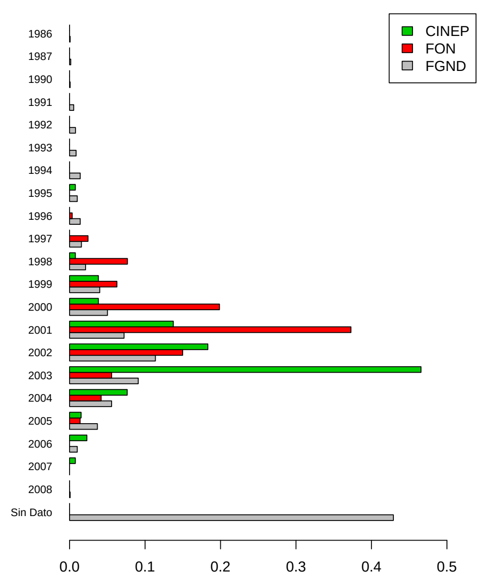
3.4.2 Proportion of Victims in each Dataset by Time and Space

Sometimes, one dataset is used to try to explain "the story" about "what happened." Below, we present a series of graphs about how these stories may subtly – or evidently – differ depending on which dataset is used. In debates, the collectors of each dataset tend to explain the differences based on definitional categories or on politically-motivated manipulations to the data. We believe each list is collected to the best ability of each organization, and that the differences are due to varying levels of access, security, trust with the victim population, and resources including staff, logistics, and information processing capacity.

Figures 12 - 17 show three datasets per graph (chosen by largest datasets we have on that act type) and present what proportion of each list has records for each year or each municipality. In other words, if we took each list separately, what story would it tell about the "most violent" year or municipality, the "least violent", and the pattern in between. Of course, the story in not necessarily about the real pattern of violence, but rather, the pattern of observations by each organization.

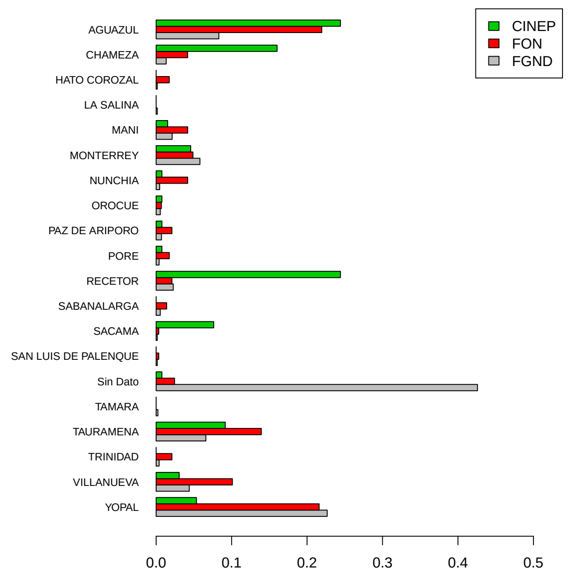
In Figure 12 we see that all three lists report a gradual rise by year and then a fall in later years. However, there is great variance by list by year. The peak observed by *Fondelibertad* (FON) is in 2001 and the peak in CINEP is in 2003. The highest proportion of disappearances reported by the *Fiscalía* (FGND) have no date information.

Figure 12: Proportions of Disappearances by Dataset by Year



When we compare the same three datasets by municipality, we see very interesting differences in the patterns. CINEP observes the highest proportions of disappearances in Aguazul and Recetor. The other two show proportionally more violations in Yopal and to a lesser extent Aguazul. Again, FGND has a very high proportion of data without municipality information. FON is capturing data that the other two have not in Hato Corozal, Nunchia and Trinidad.

Figure 13: Proportions of Disappearances by Dataset by Municipality



For killings, we present two sets of graphs. Figures 14 - 15 plot the proportions of the the largest datasets for killings, and include general homicides. Figures 16 - 17 compare datasets that only record socio-political killings.

Figure 14 makes explicit that the datasets shared with us do not cover the same years.

Figure 14: Proportions of Killings by Dataset by Year

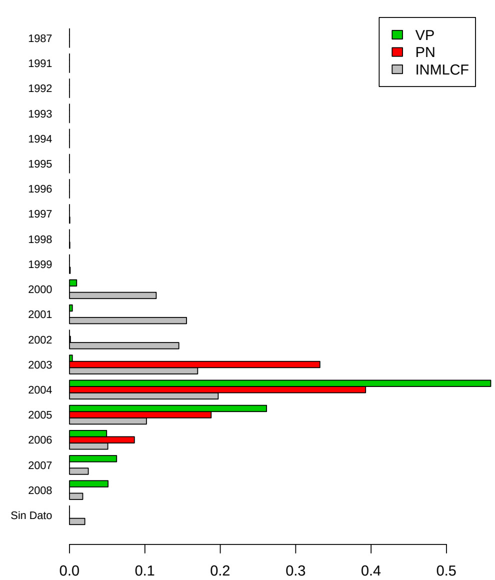
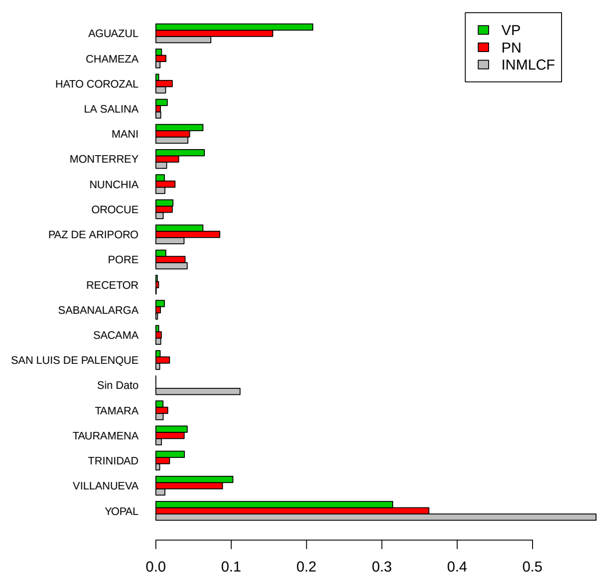


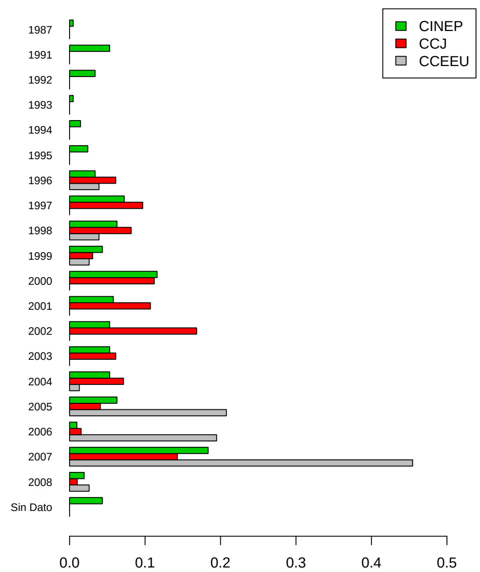
Figure 15 shows that INMLCF, PN and VP would tell a similar story by municipality, although INMLCF would tell a more dramatic story about Yopal than the other two.

Figure 15: Proportions of Killings by Dataset by Municipality



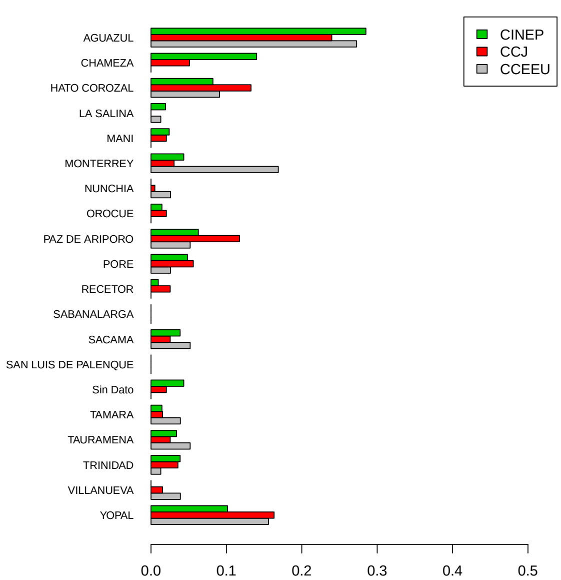
In Figure 16, we see that CINEP and the Colombia-Europe-US Coordination (CCEEU) have the highest coverage of killings in 2007. Colombian Commission of Jurists (CCJ) shows similar proportions in 2002 and 2007.

Figure 16: Proportions of Political Killings by Dataset by Year



Finally, in Figure 17, we see that all three lists show close to 25% of their reports in Aguazul. CCEEU is observing a relatively high proportion of coverage in Monterrey. CCJ is observing higher proportions in Hato Corozal and Paz de Ariporo than the other two lists.

Figure 17: Proportions of Political Killings by Dataset by Municipality



We believe these differences in reporting should be expected. The dissimilarities result from organizations facing the challenges in data collection and reporting with different mandates, different access to social networks, and different levels of resources. No one list holds "the truth." They all tell part of the story, and they all complement the creation of knowledge about past violence. Descriptive analysis is an intermediate analytic step.

To overcome the differences among individual datasets, statistical methods that make statistical inferences from multiple data sources are vital. Individual data sources are incomplete, but together they illuminate our understanding of the universe of violent acts under study. By relying on multiple sources of data, Multiple Systems Estimation (MSE) provides a way to estimate the amount of under-registration - events or individuals that may not be documented in any source - and thus ultimately estimate the total number of violent acts, both documented and undocumented.

4 Conclusion

Given the amount of uncertainty that still exists after combining data from 15 datasets, and calculating estimates of unknown and total victims, we recommend strongly against using any one dataset to make claims about trends or patterns of violence. Claims about patterns of violence are important because they often aim to impact policy, drive intervention efforts and clarify history.

We believe that estimates of lethal violence matter historically for several reasons. First, unreported lethal acts represent a disappearance in social knowledge. Undocumented victims may become entirely invisible because there is no persistent record of their death or disappearance. Some victims may become invisible even if their bodies may be recovered because the bodies are not identified; these are the no name (NN) cadavers. In some cases, the killing or disappearance may be known only to the victim's family who has chosen not to report the crime. Our estimates of the number of uncounted killings and disappearances ensure that these invisible victims are not lost from the historical understanding of violence.

5 Future Work

Although our analysis of killings and disappearances in Casanare represents a major advancement in our methodology, much more can be done. We are further developing the details of this method in the academic statistics literature. We look forward to applying this and other statistical methods in additional departments and eventually across all of Colombia.

Acknowledgement

This project was funded by the Swedish International Development Cooperation Agency (SIDA) in Colombia and by core support to the Benetech Initiative from the John D. and Catherine T. MacArthur Foundation, The Oak Foundation and the Sigrid Rausing Trust. We are very grateful for their ongoing support.

We were fortunate to work with EQUITAS in the funding, data gathering and dissemination of this work. We would also like to thank the fifteen organizations for making their data about killings and disappearances available to us or to our partner EQUITAS. We want to especially recognize the local experts on Casanare who on numerous occasions shared their qualitative knowledge about the conflict in Casanare with us and helped us develop meaningful strata.

A special thank you to David Banks at Duke University and Bala Rajaratnam at Stanford University for reviewing the methodological developments to MSE used for this study.

Finally we want to thank the Benetech Human Rights Program team members who contributed to the project, in particular: Jeff Klingner for calculating the inter-rater reliability result and for the map plots; Ann Harrison for her careful edits; and Beatriz Vejarano for her translation to Spanish of the report and for her support on many of the day-to-day tasks involved in completing this project.

About the Authors

Tamy Guberek, B.A., is the Latin America Field Coordinator for the Benetech Human Rights Program. Ms. Guberek works with partners in Colombia and Guatemala to conduct defensible data analysis and to improve their human rights information management systems. She received a dual B.A. in International Relations and Peace and Justice Studies from Tufts University.

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.

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.

Kristian Lum, M.S., is a former summer intern with the Benetech Human Rights Program and is currently a statistical consultant. She is a Ph.D. candidate in the Statistical Science Department at Duke University.

Patrick Ball, Ph.D., is the Vice President of Human Rights Programs at the Benetech Initiative. Since 1991, Dr. Ball has designed information management systems and conducted statistical analysis for large-scale human rights data projects used by truth commissions, non-governmental organizations, tribunals and United Nations missions in El Salvador, Ethiopia, Guatemala, Haiti, South Africa, Kosovo, Sierra Leone, Perú, Timor-Leste, and Chad. Dr. Ball is currently involved in Benetech projects in Colombia, Burma, Lebanon, Guatemala and in other countries around the world.

About the Benetech Human Rights Program

The Benetech Human Rights Program has more than 17 years of experience applying rigorous scientific analysis to data about human rights violations. Our expertise has been sought by nine Truth and Reconciliation Commissions, by U.N. missions and official human rights bodies, and by international criminal tribunals. We have conducted projects in El Salvador, Ethiopia, Guatemala, Haiti, South Africa, Kosovo, Sierra Leone, Sri Lanka, Timor-Leste, Colombia and Peru; and provided extensive guidance on data processing and analysis methodologies to non-governmental organizations and partner groups in many countries throughout the world. With our partners, we make scientifically-defensible arguments based in rigorous evidence (http://www.benetech.org, http://www.hrdag.org).

The materials contained herein represent the opinions of the authors and editors and should not be construed to be the view of the Benetech Initiative, any of Benetech's constituent projects, the Benetech Board of Directors or the donors to Benetech.

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

Appendix: Methodology

MSE has been refined for estimating human populations in censuses [8] [9], and the authors of this paper have used MSE to estimate total conflict-related mortality in several cases. [10] [11] [12] [13] Unfortunately, our research on violent acts in Colombia presented several challenges to the MSE methodology we had previously implemented. These challenges stemmed from a successful data gathering effort in the Colombian department of Casanare, which yielded 15 independent datasets documenting violent acts. The following two sections outline the method we previously implemented in Casanare and provide details of the current method.

Previous Implementations of MSE in Casanare

The most basic version of MSE relies on four 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.
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. There is *perfect matching*: it is possible 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.

The first assumption, that the object of measurement– whether it is a population of persons in a country or a population of violent events that occurred in a state– is a closed system, is easily satisfied by our data. This assumption is generally unproblematic for retrospective data on violent events, because events that occurred cannot "un-occur" later. Said differently, during the year 2010, no one can die in the year 2009: the population of deaths in 2009 is closed.

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. One way to address unequal capture probabilities is through stratification. It makes sense intuitively to stratify over both space and time, since both geographic areas and different periods of time are likely to have different probabilities of capture. Specific stratification details for Casanare can be found in Section 2.1. Effective stratification requires that in each stratum there be sufficient data in all systems and sufficient overlap among systems.

The third assumption, independence of systems, is similarly difficult to meet. A common example here is the difference between governmental and non-governmental organizations. Because different populations may have varying levels of trust in the two kinds of 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

[8] 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):11371148, 1993.

[9] Chandra C. Sekar and Edwards W. Deming. On a method of estimating birth and death rates and the extent of registration. *Journal of the American Statistical Association*, 44(245):101115, 1949.

[10] Patrick Ball. Political Killings in Kosova/Kosovo March-June 1999. American Bar Association, 2000.

[11] Patrick Ball. *Making the Case: Investigating Large Scale Human Rights Violations Using Information Systems and Data Analysis*, ch. 12. American Association for the Advancement of Science, Washington, DC, 1999.

[12] Patrick Ball, Jana Asher, David Sulmont, and Daniel Manrique. How Many Peruvians Have Died? an estimate of the total number of victims killed or disappeared in the armed internal conflict between 1980 and 2000. Report to the Peruvian Commission for Truth and Reconciliation (CVR), Washington, DC, August 2003.

[13] Romesh Silva and Patrick Ball. The Profile of Human Rights Violations in Timor-Leste, 1974-1999. Report by the Benetech Human Rights Program to the Commission on Reception, Truth and Reconciliation (CAVR), 2006.

other.

The fourth assumption, perfect matching between systems, leads to the most computationally intensive part of the multiple systems process. At present there exist no tractable models for MSE with imperfect matching; the task instead is to match records as accurately as possible using some unique identifier(s). See Section 3.2 for details about this process in Casanare.

To summarize, assumptions one and four are believed to be met by our data. We adjust for assumption two through stratification. In the past report on Casanare, we adjusted for assumption three by modeling the dependencies between lists using loglinear models, as suggested by Bishop et al. [14] Unfortunately, the multitude of lists available in Casanare makes this method impossible (it is computationally infeasible to enumerate every possible loglinear model following the method suggested by Bishop et al.).

Therefore new methodology had to be developed to estimate the total number of victims of lethal violence in Casanare. The following section outlines this method.

Current Implementation of MSE in Casanare

One of the first challenges to overcome is the need for data reduction. Enumerating all possible models following the loglinear approach is not possible, but we do not want to discard any data. Since the theory and methodology of MSE is well-developed for three systems (lists or datasets) we chose to look at every possible combination of three lists from our total group of 15. This results in $\binom{15}{3} = 455$ possible data partitions. In practice, there were actually fewer "valid" partitions. A partition was considered valid if the three datasets in that specific partition contained sufficient overlaps to calculate an estimate using MSE.

As mentioned in Section 2.2, for some strata there was not enough data to calculate an estimate. "Not enough data" implies one of two things: 1) every dataset only observed a small number of acts in a given stratum, and there is little to no overlap among datasets, or 2) datasets recorded many different acts, but there remains no overlap. The second scenario, many different events, indicates the possibility that the unobserved population of lethal acts is in fact quite large. However, there is no way to use the available data to generate estimates for these cases of insufficient overlap. Therefore, by omitting these cases we may be under-estimating the total number of lethal acts in Casanare, i.e., our estimates may be conservative. We believe that for the strata for which the available data is very limited, or for which the observable overlaps are limited, our estimates are consequently biased downward.

Within each partition with sufficient overlap, we can build several possible models to account for potential dependencies between lists. The next challenge is determining which model to choose. Often the Bayesian Information Criterion (BIC) is used to indicate the goodness of fit of a model and the model with the lowest BIC will be chosen as the 'best' model.

However, in many cases, there are many plausible models that produce very different estimates. The Bayesian Model Averaging (BMA) technique takes uncertainty about the model itself into account. The underlying Bayesian assumption is that each of the parameters in the model and the model itself are random variables rather than fixed values. To learn about the parameters and models, one must learn about their distribution. That is, a Bayesian would like to learn about $P(N|\mathbf{Y})$, the probability of any N (the estimated total number of observed and unobserved acts), given our observed data $\mathbf{Y}$. Each element in our observed data $\mathbf{Y}$ is a count of the total number of names to appear in each intersection of the lists. $P(N|\mathbf{Y})$ can be calculated as

$$P(N|\mathbf{Y}) = \sum_{\mathcal{M}} p(N|\mathbf{Y}, M)p(M|\mathbf{Y}), \tag{1}$$

[14] Yvonne M. M. Bishop, Stephen Fienberg, and Paul H. Holland. *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge, MA, 1975.

where $P(N|\mathbf{Y}, M)$ is the distribution of the unknown model parameters given the observed data and any one candidate model M in the set of all possible models $\mathcal{M}$ and $p(M|\mathbf{Y})$ is the posterior probability of model M.

A key piece of this equation, $p(M|\mathbf{Y})$ can be re-written as

$$p(M|\mathbf{Y}) = \frac{p(\mathbf{Y}|M)p(M)}{\sum_{\mathcal{M}} p(\mathbf{Y}|M)p(M)} \tag{2}$$

where $p(M)$ is the prior probability of model M and $p(\mathbf{Y}|\mathbf{M})$ is the marginal likelihood of the data. Unfortunately, $p(\mathbf{Y}|M)$ only has an analytical solution under special circumstances. Therefore, we define the model for our data in a way that is analytically useful.

We use a Multinomial likelihood for our data and specify a Dirichlet prior for the model parameters. Because of a property called conjugacy between the Multinomial likelihood and the Dirichlet priors, we can derive the analytical solution to $p(\mathbf{Y}|M)$.

$$\begin{aligned} Y|\theta, N, M &\sim Multinomial(\theta) \\ \theta|M &\sim Dirichlet \\ Y|N, M &\propto \int_\theta [Y|\theta, N, M][\theta|M] \end{aligned}$$

where θ is a collection of probabilities for each count in Y.

We can also assign a prior distribution to N, the total number of lethal acts in Casanare, both observed and unobserved. The prior we choose is called a truncated prior; we use such a prior not only for computational convenience, but also because we believe that the total number of lethal acts in Casanare cannot be any greater than some reasonable proportion of the entire population of Casanare. Thus, similar to what is suggested in Madigan and York[15], we use the prior $\pi(N) \propto N^{-1}1[N_* \leq N \leq N^*]$ which takes positive probability only on $[N_*, N^*]$ for N_* the total number of unique names that appear in the datasets (because we know there must be at least as many total acts as are observed) and N^* an absolute upper bound on the number of possible lethal acts in the region.[16]

Now that we have fully specified a model, we can begin to talk about the posterior distribution of N. For a given model, if we would like to know the posterior probability of any N, we need only calculate:

$$P(N|\mathbf{Y}, M) = \frac{P(Y|N, M)\pi(N)}{\sum_{N=N_*}^{N^*} P(Y|N, M)\pi(N)}$$

However, as described above, we need to average over all of our possible models.

$$P(N|\mathbf{Y}) = \frac{\sum_{M\in\mathcal{M}} P(Y|N, M)\pi(N)\pi(M)}{\sum_{N=N_*}^{N^*}\sum_{M\in\mathcal{M}} P(Y|N, M)\pi(N)\pi(M)}$$

We simply set $\pi(M) \sim |\mathcal{M}|^{-1}$, where $\mathcal{M}$ is the set of the full universe of models.

[15]Madigan, David and Jeremy C. York. 1997. "Bayesian Methods for Estimation of the Size of a Closed Population" *Biometrika*. 84(1), 19-31

[16]The absolute upper bound for analyses in this report was determined by multiplying twice the average mortality rate for all of Colombia in 2007 times the population of Casanare. This resulted in a value of N^* of 3,324. Therefore, it was assumed that 3,324 was the absolute upper limit on the possible number of killings or disappearances in a given group of municipalities within Casanare in any one year.

One last layer of complexity, as mentioned above, we do not have only three datasets; we have 15. Now let $\mathbf{Y}^\gamma$ be the γth combination of three datasets. Here, we let $\mathcal{M}(\gamma)$ represent all the model representations of the three datasets indexed by γ. Now, averaging over each of these combinations, all that remains is to re-assemble the pieces similar to before:

$$\begin{aligned} P(N|Y) &= Z^{-1} \sum_{\gamma} \sum_{M \in \mathcal{M}(\gamma)} P(Y^\gamma|N, M)\pi(N)\pi(M, \gamma) \\ Z &= \sum_{N=N_*}^{N^*} \sum_{\gamma} \sum_{M \in \mathcal{M}(\gamma)} P(Y^\gamma|N, M)\pi(N)\pi(M, \gamma) \end{aligned}$$

Results from applying this method to the data from Casanare are presented in Section 2.

One of the key differences in terms of interpreting results from this method as compared to our previous method in the 2007 report is the way uncertainty is measured. Fully Bayesian methods, such as those outlined in this section, result in estimates of 95% *credible* intervals. These differ from 95% *confidence* intervals. Specifically, credible intervals are based on a fully specified distribution, which means we are 95% sure that the correct number of violent acts is within the interval (given that the interval is based on our observed data). This is a much more straight-forward interpretation of uncertainty as compared to traditional confidence intervals. Traditional confidence intervals tell us that if we conducted the study 100 times, 95 of these times we would expect the answer to fall in the reported range.