Missing People in Casanare

Daniel Guzmán, Tamy Guberek, Amelia Hoover, and Patrick Ball

November 28, 2007

1 Introduction

How many people are missing in the department of Casanare, Colombia? This apparently simple question proves complex when we ask how many missing persons were not reported to any organization, and becomes even more difficult in the context of politically contentious debates about exhumation, identification and reunification of remains. How can we be sure that all the missing are accounted for in some way? How should we approach the problem of searching for victims? Answers to these and other questions will be incorrect if we assume that any list or combination of lists is "comprehensive." Ultimately, correct answers rely on scientific estimation of the number of missing persons.

In this initial analysis, we estimate that the total number of missing persons in Casanare 1986-2007 is 2,553.[1] Approximately 1,500 persons were *reported* missing during this period, yielding an "undocumented rate" of about 40% (of total estimated missing persons). We emphasize that the rate of undocumented missing persons found in Casanare does not necessarily represent the rate that could be found in Colombia more generally, if data were available. We recommend that additional data should be gathered and made available for analysis by statisticians and social scientists. Furthermore, this analysis demonstrates that no single list of missing persons estimates the total number of persons likely to be missing in Casanare.

We proceed in several sections. First, we outline our findings. Then we consider the available data on Casanare from thirteen data collection projects. A section on under-registration describes complexities related to data that are *not* available. Then we describe our data and our matching and estimation techniques in more detail. Finally, we lay out a research agenda designed to provide a rigorous and data-driven picture of violence in Casanare and in Colombia more generally.

2 Findings

Our analysis suggests that between thirty and forty percent of missing persons in the Department of Casanare went unreported during 1986-2007. Analyzing the

[1] The estimated 2553 is in a 95% confidence interval 2239-2867; see Table 1

overlaps between 1,524 unique missing persons reports from three data systems yielded an estimated 1,029 unregistered missing persons, for a total of 1,524 + 1,029 = 2,553 missing persons. These estimates, and the level of uncertainty attached to the estimates, are discussed below. In addition to our estimates of the total number of unreported missing persons in the department, we also estimated unregistered missing persons for 7 subsets of the data.

Our findings are summarized graphically in the following figure:

Graph 1: Estimates of Total Missing Persons in Casanare 1986-2007

The same results are presented in the table, below.

Table 1: Estimates of Total Missing Persons in Casanare 1986-2007

Stratum	CI low	Estimate	CI high	Percent Undocumented
Tauramena Aguazul Yopal 1998-2005	607	803	999	38%
Tauramena Aguazul Yopal 1986-2007	794	1049	1304	40%
South 2001-2004	648	746	844	30%
South 1998-2005	978	1119	1260	32%
South 1986-2007	1220	1392	1564	34%
Casanare 2001-2004	713	821	929	31%
Casanare 1998-2005	1088	1239	1390	34%
Casanare 1986-2007	2239	2553	2867	40%

It is important to emphasize that these are *estimates*, and that every estimate is associated with a measure of uncertainty. We express our degree of uncertainty by reporting our estimates with their associated “confidence intervals” in the form Z (X, Y), with a lower bound (X) and upper bound (Y) centered at the estimate (Z). In the case of our estimates, the confidence interval should be interpreted as a statement that there is a 95% likelihood that the true number of unregistered missing persons lies between the lower and upper bound. (For a more technical discussion of the construction of our estimates and their associated confidence intervals, see below at section 7.)

For the department as a whole between 1986 and 2007 (represented by the dark blue area in the graphic), we estimate 1029 (715, 1343) unregistered missing persons. Adding this estimate to the number of already registered missing persons, 1524, we estimate 2553 (2239, 2867) total missing persons. We also created estimates for two subsets of the full-department data, over the time periods 1998-2005 (lighter blue) and 2001-2004 (very light blue). The estimated number of unregistered missing persons in the department as a whole between 1998 and 2005 is 416 (265, 567), which is added to 823 registered missing persons for an estimated 1239 (1088, 1390) total missing persons. For the still-shorter time period 2001-2004, we estimate 257 (149, 365) unregistered missing persons, which is added to 564 registered missing persons for an estimated 821 (713, 929) total missing persons. An important logical test for these estimates is that shorter time periods should have lower estimates than longer time periods, and indeed $821 < 1239 < 2553$.

We estimate unregistered persons for two regional strata as well. One stratum includes the southern region of Casanare (municipalities Sabanalarga, Villanueva, Monterrey, Aguazul, Tauramena, Maní, Chameza, Recetor and Yopal), represented above as "SOUTH" in shades of grey. In SOUTH, we estimate 479 (307, 651) unregistered missing persons between 1986 and 2007, which is added to 913 registered missing persons for an estimated 1392 (1220, 1564) total missing persons. From 1998-2005, we estimate 359 (218, 500) unregistered missing persons in SOUTH, which is added to 760 registered missing persons for an estimated 1119 (978, 1260) total missing persons. For 2001-2004, we estimate 223 (125, 321) unregistered missing persons, which is added to 523 registered missing persons for an estimated 746 (648, 844) total missing persons. As with the full department estimates, the nested time subsets for the southern region including Yopal have the expected relation: $746 < 1119 < 1392$.

Our third regional subset, represented above as "TAU, AGUA, YOP" in shades of green, includes the municipalities of Tauramena, Aguazul, and Yopal (TAY). Because TAY is a relatively small subset of municipalities, there was insufficient data to calculate an estimate for 2001-2004 in TAY. However, we estimated unregistered missing persons in this subset for 1986-2007 and 1998-2005. For 1987-2006, we estimate 422 (167, 677) unregistered missing persons, which is added to 627 registered missing persons for an estimated 1049 (794, 1304) total missing persons. For 1998-2005, we estimate 303 (107, 499) unregistered missing persons, which is added to 500 registered missing persons for a estimated 803 (607, 999) total missing persons. Again, logically we should find that the estimate for 1998-2005 is lower than that for 1986-2007, and indeed it is: $803 < 1049$.

Importantly, estimates from regional subsets also nest as they should across time. TAY is a subset of SOUTH, which is a subset of the full department. Consequently, we should expect that within each time stratum, the regional estimates will nest appropriately (TAY < SOUTH < CASANARE). And indeed this is the case. For the full time span, 1986-2007, we estimate 1049 total missing persons in TAY, 1392 total missing persons in SOUTH and 2553 total missing persons across Casanare. For the 1998-2005 subset, we estimate 803 total missing persons in TAY, 1119 total missing persons in SUR and 1239 total missing persons in the full department. For the 2001-2004 subset, we cannot make an estimate for TAY, but we estimate 746 total missing persons in SUR, as compared to 821 in the department as a whole.

The fact that the nested time and region estimates "behave" as they logically should increases our initial confidence in these estimates. It is also important to note that estimated rates of underregistration (measured as the proportion of the estimated number of unregistered cases to the estimated number of total cases) is strikingly consistent across strata, varying only from 30% (estimated underregistration rate in SOUTH, 2001-2004) to 40% (estimated underregistration rate in the full department across 1986-2007 and in TAY 1986-2007).

3 Complexity in Documenting Missing People

3.1 Different and Contradictory Definitions

A missing person can often be hidden behind a spectrum of different terms. In this study, a missing person may have been reported as a disappearance, a simple kidnapping, an extortion-related kidnapping, a freed kidnapping, a death, or a hostage. The fact that a person went missing often gets filtered through institutional mandates and legal or political interpretations before its entered into a database. The level of detail given by a deponent about the event may also influence how a violation is recorded.

Furthermore, the same person reported to different projects may have been recorded as a victim of a slightly different crime. For example, if a person was abducted by an unknown perpetrator for unknown reasons and never heard from again, one organization may call this event a simple kidnapping or a disappearance. If the perpetrator was a combatant, the event may have been called a prisoner of war (in Spanish, "retenido en combate"). Alternatively, a person reported as disappeared by one group may have subsequently been reported dead by another.

Its interesting to note that in this study:

- 49 people recorded by two or more datasets were listed as a "simple kidnapping" by one organization and a "disappearance" by another;
- 46 people were said to have been victims of "extortion-related kidnappings" by one group and a "disappearance" by another;
- 4 people were listed as a "hostage" by one organization and a "disappearance" by another;

- 33 people reported as "disappeared" or categorized under "simple kidnapping" by one organization were reported dead by another.

We have made the attempt to capture a broad universe of missing people, even where there were contradictions in the way cases were reported. If the fate of the victim was eventually reported – if the person was freed or their body was found – either within the same dataset or by different one, we considered them no longer missing and dropped those records from the study.

3.2 How Faulty Can Data Be?

The problem with data is not only different and contradictory definitions. Here we outline other common problems with human rights violation data that must be considered and corrected before proceeding to data analysis.

3.2.1 Missing Data

While we may have information that a person went missing, we do not always know much more than that. For disappearances in particular, often times the perpetrator invests great effort and resources into not only disappearing a person, but also concealing as much detail about the occurrence as possible. Sooner or later, it becomes obvious that a person is missing, but details about when, where, why, how or by who may be unavailable. Approximately 1/4 of the data used in this study did not have the municipality nor the date when the disappearances took place. Records that lacked information about the time or place the person went missing were used for the global analysis, but they were excluded from the estimates stratified by time and space.

3.2.2 Duplicate Reports

Duplicate records (more than one record referring to the same event) present a major problem for any organization collecting data on violent events. Duplicates occur for many reasons, both across datasets (for example, when a witness to an event makes several reports to several organizations) and within datasets (for example, when multiple witnesses to the same event report that event to one organization). Duplicates can also occur when organizations collect and combine lists from other sources.

As we explain below in section 7, *recognized* duplicates across datasets are necessary for estimating the number of uncounted violent events using multiple systems estimation. *Unrecognized* duplicates, on the other hand, can lead to overcounting of violent events. For example, data from the Fiscalía General de la Nación included nearly 1,500 recorded missing persons in Casanare between 1986 and 2007. However, some of these records were duplicates; the same disappearances were included in the data multiple times. After accounting for duplicates, the Fiscalía data included only 1331 unique cases of disappearance. Other datasets included in this analysis also included significant numbers of duplicates, as shown in the table below.

Table 2: Records and Match Groups in 13 Datasets

Dataset	Records	Unique Persons
Registro Ünico de Cadaveres	1795	1786
Fiscalía General de la Nación	1482	1331
Policía Nacional	828	827
Fondelibertad	711	702
Comisión Colombiana de Juristas	384	380
Gaula	359	356
Fiscalía Santa Rosa de Viterbo	279	277
Registro Único de Desaparición	203	190
Instituto Nacional de Medicina Legal	112	112
Cuerpo Técnico de Investigación	109	109
Familiares Colombia	46	45
Fundación País Libre	39	38
ASFADDES	6	6

Duplicate cases can cause serious errors in conclusions about the pattern and magnitude of violence. Duplicates often do not lead to inflated estimates of the total number of violent events, because many cases are never counted at all. But imagine that certain violent events (perhaps missing persons in municipality X) are very likely to be reported multiple times, whereas other violent events (perhaps missing persons in municipality Y) are less likely to be duplicated – or unlikely to be reported at all. If this is the case, we will overestimate the number of disappearances in municipality X, *and* X will appear far more violent, relative to Y, than it actually is.

In this analysis, we identified duplicate cases by matching records against one another (see Section 6 below), identifying clusters of records that were likely to refer to the same event.

4 Under-registration and Bias

We can think of uncounted cases of violence ("underregistration") as the opposite of duplicate cases: whereas duplication could lead to overestimates, underregistration leads to underestimates of the number of missing persons. Likewise, where duplications can lead to artificial inflation of the relative number of cases from one area or group, systematic underregistration can artificially deflate the relative number of cases from one area or group. Artificial inflation or deflation of one group or area, relative to another, causes errors in conclusions about patterns of violence. However, correcting for underregistration is a more difficult problem than identifying duplicates. We can directly observe the number of duplicate cases in the data (by looking for matching records), but we can only estimate the number of uncounted cases. This estimation is a key focus of this report (see Section 7 below on methodology), because failing to account for underregistration strongly affects the conclusions reached about both the pattern and the magnitude of violence.

Considering common reasons for underregistration allows us to anticipate the biases that we might expect in a given dataset, and can help explain some differences between datasets. For example, geography and accessibility issues can strongly influence reporting. Suppose that two organizations collect testimonies about violence in a department. One organization has offices in a remote village, while another organization has offices only in the capital city. Then it is likely that violent events in the village will be underreported by the organization based in the city, and that violent events in the capital will be underreported by the village office. If violence decreases in the capital but not in the village, the organization based in the capital might report that violence is declining in the state, while the village-based organization would reach an entirely different conclusion.

A similar effect could occur because of potential respondents' social networks, or their confidence in particular institutions. Perhaps one group of potential respondents is very unlikely to report episodes of violence to the police, but very likely to report to a non-govermental organization. In this case, we might expect that if this group of potential respondents were targeted for increasing violence, the trend would appear in the NGO data but not in police data.

If we could assume that the rate of reporting never varied, or that different groups of violations were reported at roughly the same rate, it might make sense to draw conclusions about trends in violence from single datasets. But such assumptions never hold. However, even before formally estimating the rate of under-registration, we can think through likely sources of under-registration and the biases that might result. For example, organizations based in particular areas are likely to be "biased toward" those areas, in that they report cases of violence there more completely. Organizations that have a particular constituency (for example, the church, or labor unions) are likely to be biased toward those constituencies and to under-report cases from other constituencies.

5 Data

In order to estimate the total universe of missing people in Casanare, we used data collected by 13 organizations. As mentioned earlier, the data used for this study included all records in Casanare classified as either a disappearance, a simple kidnapping, an extortion-related kidnapping, a freed kidnapping, a death, or a hostage. The total number of records with these definitions equaled 6,353.

Definitions outside the scope of the definition for "missing person" were included at the start so that if these records matched a disappearance or simple kidnapping, both records could be dropped from the study. In other words, if Jane Doe was said to be disappeared in dataset 1 and Jane Doe was said to be dead in dataset 2, we dropped the record for Jan Doe from *both* datasets. If we left Jane Doe in dataset 1 and dropped her from dataset 2, it would seem like these two datasets did not overlap, and therefore the our estimate of the undocumented would increase. Dropping Jane Doe from both datasets is a conservative decision so we minimize the risk of over-estimating the missing person universe.

The following records were dropped because they did not meet the definition of "missing":

- all deaths
- all disappearances and simple kidnappings that matched a death
- all extortion-related and liberated kidnappings, and any other records that matched these violations.
- all hostages that did not match a disappearance or simple kidnapping

After this round, 2,138 records and 1544 unique individuals were left in the study[2].

In addition to the data that were dropped due to definitional ineligibility, there were also data not sufficient to be included in the estimates. Approximately 25% of the data that we have had to be partially excluded from stratified estimates because they lacked the key elements used for these calculations, namely date and municipality of disappearance.

We used data from the department of Casanare. Table 2 above lists the 13 organizations that generously shared their data with EQUITAS, Benetech, and other partners for this project.

6 Matching

All the data from the 13 datasets were standardized into one list. The list was sorted by different variables in an effort to group records that referred to the same person. A unique identification number was assigned to each match group of records that, according to our criteria, represented the same person.

To be considered eligible for a match group, records had to have at least two names, and up to four names[3], place and date of disappearance. They had to match on most of these variables and could not contradict on sex. The decision about whether two or more records referred to the same person was made by members of HRDAG.

[2] Note that there is a difference between the number of unique individuals in the study as a whole and the number of unique individuals used in our estimates of unregistered missing persons. 1,544 unique missing persons, grouped into four overlapping systems, are considered in the study as a whole. However, our statistical model employs three of these four systems, including 1,524 unique missing persons, to calculate the estimated number of unregistered missing persons. For more information on model selection criteria, see below at Section 7.2.3, Model Selection.

[3] First name, middle name, father's last name, and mother's last name.

6.1 Matching Consistency

To ensure that matching decisions were made consistently, two different HRDAG members independently matched some of the same records. We measured the consistency of their matching using a metric known as the *F-measure*. This measure combines the rate at which one matcher finds matches also found by the other (recall) with how often the first matchers' matches are found by the second (precision). If m_1 is the number of matches found only by the first matcher, m_2 is the number found only by the second, and m_b is the number found by both of them, then

$$precision = \frac{m_b}{m_1 + m_b} \quad (1)$$

$$recall = \frac{m_b}{m_2 + m_b} \quad (2)$$

$$F\text{-}measure = \frac{2 \times precision \times recall}{precision + recall} \quad (3)$$

Our two matchers had an average F-measure of 0.72 for finding duplicates within each dataset and an F-measure of 0.96 for finding matches between different datasets. More specifically, this means that for pairs of records from different datasets, 691 record pairs were identified as matches by both matchers, 16 record pairs were matched only by the first matcher, and 39 record pairs were matched only by the second matcher.

7 Methods

This section describes in detail the methods we used in our Casanare analysis. In section 7.1, we introduce multiple systems estimation (MSE), showing how estimates are derived for any problem in which overlaps are used to determine the extent of under-registration. In section 7.2, we describe in detail how MSE was applied in the case of Casanare.

7.1 Multiple Systems Estimation

In order to estimate the extent of under-registration in Casanare (i.e., the number of missing persons who were not counted by any list), we use a technique known as multiple systems estimation (MSE). MSE has been refined for estimating human populations in censuses;[4] [5] the authors of this report have used

[4] Darroch, John, Stephen Fienberg, Gary Glonek, and Brian Junker. 1993. "A three-sample multiple-recapture approach to census population estimation with heterogeneous catchability." *Journal of the American Statistical Association* 88(423): 1137-1148.

[5] Chandra Sekar, C. and W. Edwards Deming. 1949. "On a method of estimating birth and death rates and the extent of registration." *Journal of the American Statistical Association* 44(245): 101-115.

MSE to estimate total mortality in several cases.[6] [7] [8] [9]

MSE is derived from probability theory: with two random samples from a closed population of unknown size, first determine how many cases were counted in both samples (this is called the overlap). Given the sample sizes and the size of the overlap, the size of the unknown population can be estimated (see below at Two-system estimates). In the case of missing persons or other violent events, the "unknown population" is the number of violent events that actually occurred, and the "samples" are lists of known violent events. However, estimates based on calculations from two lists rely on assumptions that data on violence can rarely meet, if ever. In this section, we first describe how simple estimates are derived for two systems. Then we describe the problems with two-system estimates, and show how these issues can be resolved by using three or more systems. In the following section (7.2), we describe more specifically how we implemented multiple systems estimation in the case of Casanare.

7.1.1 Two-system estimates

Above in Section 5, we discussed how underregistration in single datasets may cause errors in our conclusions about patterns of violence. Data on missing persons relies on the ability of organizations to obtain information about missing persons. Because reports about some missing persons are easier to find than reports about others, datasets will be *biased* toward those cases. Organizations' abilities to obtain information may also change over time. In the theory of measurement, "reliability" refers to the ability to obtain the same (or very similar) results from repeated measurements of the same object – in this case, the number of persons going missing in Casanare. Because organizations may report a different proportion of the total missing persons in each period of time, single datasets are likely to be *unreliable*.

Under certain special circumstances, two systems can provide a more reliable and less biased result than a single dataset (Chandra Sekar and Deming 1949).

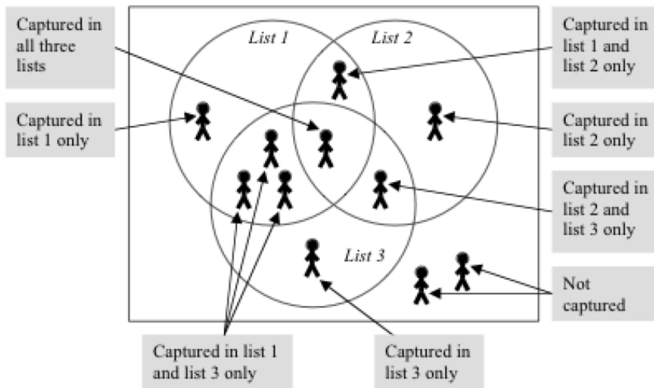
---

[6] Ball, Patrick. 1999. "Metodología intermuestra." Guatemala: Memoria del Silencio. Vol. 12. CEH. Reproduced in English in Patrick Ball, Herbert Spirer, and Louise Spirer, eds. 2000. *Making the Case: Investigating Large Scale Human Rights Violations Using Information Systems and Data Analysis.* Washington, DC: AAAS.

[7] Ball, Patrick (with the American Bar Association-Central and East European Law Initiative). 2000. "Political Killings in Kosova/Kosovo, March-June 1999." Washington, DC: ABA/CEELI-AAAS.

[8] Ball, Patrick, Jana Asher, David Sulmont, and Daniel Manrique. 2003. "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 Justice (CVR). Also published as *Anexo 2 (Anexo Estadístico)* of CVR Report, 28 August 2003. Washington, DC: AAAS.

[9] Silva, Romesh, and Patrick Ball. 2006. "The Profile of Human Rights Violations in Timor-Leste, 1974-1999." Report by the Benetech Human Rights Data Analysis Group to the Commission on Reception, Truth and Reconciliation (CAVR). Also published as Part 6 of *Chega! Final Report of the Commission for Reception, Truth and Reconciliation in East Timor.* Available online `http://hrdag.org/resources/timor_chapter_graphs/timor_chapter_page_01.shtml`



If every individual in a population of size N is equally likely to be sampled in sample A, the probability of "capture" for a single individual is(sample size A)/(population size N). If another sample, B, is also taken from the same population, the probability of capture in sample B is (sample size 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 the probability of being sampled in both A and B (call that group M) is just ((sample size A)/(population size N)) $\times$ ((sample size B)/(population size N)). But note that the probability of being in group M is *also* equal to (size of M)/(population size N).

Numerically, that is

$$Pr(A) = \frac{A}{N} \tag{4}$$

$$Pr(B) = \frac{B}{N} \tag{5}$$

$$Pr(M) = \frac{M}{N} \tag{6}$$

and

$$Pr(M) = Pr(A \, and \, B) = P(A)P(B) = \frac{AB}{N^2} \tag{7}$$

If we know A, B and M, then we can derive (an estimate of) the unknown population size, N.

$$\frac{M}{N} = \frac{AB}{N^2} \tag{8}$$

$$MN^2 = ABN \tag{9}$$

$$\hat{N} = AB/M. \tag{10}$$

Equation (10) is the two-system estimator for the unknown population N. However, the assumptions required for the two-system approach are very strong, as we see from equations (4)-(7). First, "N" must refer to the same population in each of (4), (5) and (6); it must be a *closed system.* Second, if (4), (5) and (6) are to hold, it must be true that each individual in N has an equal probability of capture, i.e., the *units are homogeneous.* Third, the equality in (7) relies on the *independence of systems* A and B. If the probability of capture in A raises or lowers the probability of capture in B (or vice versa) for any individual, then (7) does not hold. Finally, we must accurately partition all the "captured" individuals into A, B, or M (where $M = A\,and\,B$). (Otherwise $Pr(M) \neq Pr(A\,and\,B)$.) The practical implication of this requirement is *perfect matching*; all records referring to the same unit must be recognizable as such.

Because of the strong assumptions outlined above, two systems are generally insufficient to correct the biases and unreliability of single datasets. 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). Two systems eliminate the completeness assumption required in order to use a single dataset (i.e., to use the single system estimator, assuming that $\hat{N} = A$): with two systems we need not assume that every missing person is counted. However, the four other assumptions described above (closed system, homogeneity of capture probability, independence of systems, and perfect matching) are still required for two systems.

The first assumption is 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: the target population does not change during the period of measurement. This assumption is generally unproblematic for data on violent events, because events that occurred cannot "un-occur" later. The fact that some missing persons are later liberated does not change the fact that they went missing in the first place.

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 go missing *and* less likely to be reported missing; 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. Like differences in capture probability, dependences between systems are impossible to account for in the 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 lists are negatively correlated (like our example of a government and a non-governmental organization), two-system estimates will be inflated. If two lists are positively correlated (perhaps two different government lists), the estimates will be deflated.

The fourth assumption, perfect matching between systems, is 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). We describe our matching process above in section 6. Below, in Section 7.1.2, 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.

7.1.2 Estimates with three or more systems

Several researchers have developed techniques to correct for unequal probability of capture (violations of assumption two) and list dependences (violation of assumption three). These corrections are useful when three or more samples (datasets) are available.[10] [11](See also Chandra Sekar and Deming 1949, Darroch et al. 1993.)

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. More theoretically, sample variance decreases with sample size (Chandra Sekar and Deming 1949), meaning that the variation in probability of capture is smaller *by definition* when the data are partitioned into strata.

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 $x_{111} = 0$).

The third MSE assumption 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.[12] [13] (Also see, e.g.,

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

[11]Fienberg, Stephen, Matthew Johnson and Brian Junker. 1999. "Classical multilevel and Bayesian approaches to population size estimation using multiple lists." *Journal of the Royal Statistical Society* 162(3): 383-405.

[12]Agresti, Alan. 1994. "Simple capture-recapture models permitting unequal catchability and variable sampling effort." *Biometrics* 50(2): 494-500.

[13]Zwane, Eugene and Peter van der Heijden. 2007. "Analysing capture-recapture data when

Darroch et al. 1993; Fienberg et al. 1999.)

A more tractable solution – the one employed in this report – is to account for unequal probability of capture, to the extent possible, using stratification, and then to model residual list dependences using the log-linear model formalized by Bishop, Fienberg, and Holland (1975). For three lists, the basic problem is estimation of the missing cell in a $2 \times 2 \times 2$ table where each cell value x describes the number of observations captured by a unique combination of the three lists. $x_{010} = n$, for example, means that n observations were counted in the second list only. Similarly, the cell value x_{111} refers to the number of units enumerated on all three lists. For three lists, eight log-linear models are possible. Where m_{ijk} is the expected cell count, one model suggests independence of the lists:

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} \tag{11}$$

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

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

Three further models account for dependence between two pairs of samples; they are analogous to

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

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

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{23(jk)} + u_{13(ik)} \tag{14}$$

Several rules of thumb have been suggested for choosing the most appropriate model. By definition, the "fully saturated" model (equation 14 above) fits the data perfectly because its seven terms are precisely equal to the number of known cells. However, it cannot be used for out of sample predictions. On the other hand, more parsimonious models (i.e., models with fewer terms) are more likely to be useful for out-of-sample prediction, but these reduced models necessarily fit the data less well.

The Bayesian Information Coefficient (BIC) balances goodness-of-fit and parsimony. 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.[14] [15] [16] 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

some variables of heterogeneous catchability are not collected or asked in all registrations." *Statistics in Medicine 26*: 1069-1089.

[14]Raftery, Adrian E. 1995. "Bayesian Model Selection in Social Research." *Sociolgical Methodology* 25: 111-163.

[15]Raftery, Adrian E. 1996. "Approximate Bayes factors and accounting for model uncertainty in generalised linear models." *Biometrika* 83:2, 251-266.

[16]Hoeting, Jennifer A., David Madigan, Adrian Raftery and Chris Volinsky. 1999. "Bayesian Model Averaging: A Tutorial." *Statistical Science* 14:4, 382-417.

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.

Regardless of the rule of thumb used to select a single model from the universe of possible models, there will be some uncertainty about model choice. Perhaps, for example, the model with the lowest BIC is only slightly lower than the next best model. To account for the uncertainty inherent in model selection, the Bayesian Model Averaging (BMA) method (Hoeting et al., 1999) combines many models, computing a weighted average across each potential model in order to create a model that (1) uses information from all the models, and (2) captures model choice uncertainty in the estimated error. While Bayesian model averaging relies on the Bayes factor, which is extremely difficult to compute, Raftery (1995) has demonstrated that the Bayesian Information Coefficient approximates the Bayes factor very closely. Moreover, the estimates derived from Bayesian Model Averaging will closely approximate estimates from a single log-linear model if there is sufficient data.

Variance estimators calculated in the BMA approach (Raftery 1995) can be large relative to those derived for single log-linear models (such as those proposed by Bishop, Fienberg and Holland 1975); this is because, as we noted above, BMA explicitly includes uncertainty about model choice in its calculations. Variance estimators express our level of certainty in the model and are typically expressed by using a 95% confidence interval. For single log-linear models, we interpret a confidence interval this way: if we repeated the analysis hundreds of times, in about 95% of the repetitions, an identically computed confidence interval would contain our estimate. Bayesian model averaging uses a more straightforward (but practically very similar) interpretation: given our data and our prior beliefs about the distribution of the data, the likelihood that this interval contains the true value is 95%. In practice, BMA confidence intervals are only slightly wider than the confidence intervals around single log-linear models.

7.2 MSE Estimates in Casanare

The numbers of overlapping records are shown below:

Table 2: Overlapping Records among Four Systems

	Security Forensic	yes yes	yes no	no yes	no no
Civil Society	Judicial				
yes	yes	1	18	32	19
yes	no	0	10	1	45
no	yes	4	106	123	1069
no	no	0	96	20	n.a.

There were 1544 total match groups, representing 1544 known missing persons, found among the four systems.[17]

[17] Note that there is a difference between (1) the number of unique individuals in the study

7.2.1 The Logic of Four Systems

After ineligible data was dropped from the study, the remaining 11 datasets with missing persons data were grouped into four systems. By reviewing the functions of all the organizations that collected the data, we noted that the groups fall into four categories: Security, Judicial, Forensic, and Civil Society organizations.

The classification of organizations into these categories was consistent with the relationships among the groups: some groups include data from other datasets. For example, Fondelibertad's includes data from the other organization in the security category. Separating Fondelibertad from the other security-category dataset would imply that the datasets within a category are independent of each other, which would be inaccurate. The four categories are the following:

Table 3: Four Systems

Security	Forensic	Judicial	Civil Society
Gaula	Instituto Nacional de Medicina Legal	Fiscalía General de la Nación	Comisión Colombiana de Juristas
Fondelibertad	Registros Único de Desaparición	Fiscalía Santa Rosa de Viterbo	Fundación País Libre
		Cuerpo Técnico de Investigación	Familiares Colombia
			Asociación de Familiares de Detenidos Desaparecidos

7.2.2 Stratification

In statistical terms, stratification means to divide the universe into smaller sub-components (called strata). Stratification is usually done on the basis of characteristics of interest in the population being observed. In this study, we stratified on year and groups of municipalities in Casanare.

Before being able to stratify, we had to resolve contradictory information for municipality and date within match groups. A person may be called "disappeared" in one dataset and "dead" in another different dataset. Similarly, other information describing the event may be reported slightly differently to two or more organizations. The date or place of the disappearance may vary due to imprecisely reported information or typographical errors during data entry.

as a whole and (2) the number of unique individuals used in our estimates of unregistered missing persons. 1,544 unique missing persons, grouped into four overlapping systems, are considered in the study as a whole. However, our statistical model employs three of these four systems, including 1,524 unique missing persons, to calculate the estimated number of unregistered missing persons. For more information on model selection criteria, see below at Section 7.2.3, Model Selection.

For the purposes of stratification in this study, we created a hierarchy of rules to resolve contradictions within match groups, as follows:

Where records differed on municipality within a match group:

1. If one record had a missing value for municipality and the other had a valid municipality, we chose the valid municipality;

2. If there were two or more municipalities, we chose the one most frequently reported;

3. If one record had the capital, Yopal, for municipality and the other had a different municipality, we chose the non-Yopal municipality;[18]

4. If two records had different municipalities, and neither was Yopal, we chose the municipality by dataset, preferring them in this order: Fiscalía General de la Nación, Registros Único de Desaparición, Registro Único de Cadaveres, Instituto Nacional de Medicina Legal, Fondelibertad, Policía Nacional, Cuerpo Técnico de Investigación, Gaula, Asociación de Familiares de Detenidos Desaparecidos, Familiares Colombia and Fundación País Libre.

5. Where records differed on year within a match group:

6. If there were two or more years reported, we chose the one most fequently reported.

7. Within the most common years, we chose the latest one.

The estimates were made with the merged values for violations, municipality, and year.

7.2.3 Model Selection

After determining potential strata, we must select the correct weighted average of models (if any) for estimating the number of uncounted missing persons in each stratum. There are four potential list combinations (including three lists in each combination) for each stratum: [Security, Judicial and Forensic]; [Security, Judicial and Civil Society]; [Security, Forensic and Civil Society]; and [Judicial, Forensic and Civil Society]. We used the following process to maximize the use of the information available in the stratum.

For each of the four potential list combinations, we calculated a conventional estimate[19], a BMA across all the models, and 95% confidence interval from the BMA combination of the BFH variance estimates. Recall that BMA models are weighted averages of individual log-linear models, and that the BIC measures

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

[19] By "conventional," we mean estimates made using the estimators defined in Bishop, Fienberg, and Holland (1975). These estimates are referred to below as BFH.

model quality for each component model of a set of nested, hierarchical log-linear models. If *no* component model for any of the four list combinations had a $BIC < 0$, then we discarded the stratum, because no useful prediction could be made. If *exactly one* BMA model contained a model for which $BIC < 0$, we selected that model. If *more than one* of the four list combinations used a component model with $BIC < 0$, then we chose the model that used the most data, where the amount of data is measured as the total number of missing persons observed in the list combination.

For example, if in a given stratum, the list combinations [S, J, F] and [J, F, C] both contained models for which $BIC < 0$, but the combination [S, J, F] contained 400 observations while [J, F, C] contained only 300, we chose [S, J, F] on the grounds that this list contained more of the information available in the stratum.

In addition to (i) developing and implementing the criteria described above for BMA model selection, we conducted the following tests in each stratum: we (ii) determined the single-model MSE estimate (of 28 potential models) with minimum BIC; (iii) calculated an absolute relative ratio of the difference between the BMA estimate and the best single-model estimate for each list combination, defined as the minimum absolute relative difference of the two estimates $\| \frac{BFH-BMA}{BMA} \|$, selecting the list combination that minimized this ratio; and (iv) noted which list combination maximized the amount of data used. In every stratum for which BMA results were available, the BMA model choice procedure described above produced results equivalent to the results of the other three tests.

8 Future Research

The first step in the study of missing persons in Casanare, or episodes of violence in any region, is to use multiple, independent datasets to estimate the true patterns of violence.[20] Given the theoretical flaws affecting analysis based on any single dataset, we have shown (Ball et al. 2007) that it is inappropriate to draw conclusions about the pattern, trend or magnitude of violence from any single dataset. But even using multiple systems, the conclusions we can currently make are limited: although we can demonstrate that there exist many missing persons in Casanare not counted by any list, this analysis is ongoing. Our future research agenda includes more specific, data-rich estimates on missing persons in Casanare and, eventually, expansion to homicides in Casanare and homicides and disappearances across Colombia.

Our most immediate priority is to gather more data on missing persons in Casanare, especially the northern region of the department, where in this phase of the project data are too sparse to make quality estimates of the number of unreported missing persons. But recall from Section 5: the line between

[20] Ball, Patrick. "Making the Case: The Role of Statistics in Human Rights Reporting." *Statistical Journal of the United Nations.* ECE 18. 2001. 163-73. See also Patrick Ball, *Who Did What to Whom?*" 2nd ed. HRDAG working paper, 2007. Forthcoming from Benetech.

deaths and disappearances can be blurry. For this reason, our second priority in Casanare is an estimate of unregistered homicides in the department. This analysis will complement our work on disappearances and provide more information about the circumstances under which both types of violations occur and are reported.

Analysis in Casanare is immensely valuable in its own right. But it also will serve as a guide to an investigation of homicides and missing persons across Colombia. While the dynamic in Casanare has been such that rigorous estimation of missing persons precedes analysis of homicide data, outside Casanare our intent is to move forward with an estimation of unregistered homicides, and then to refine our results with an analysis of missing persons.

With respect to both departmental and national analyses of violence, we conclude that renewed data collection efforts, by more organizations in more locations, is vital if we are to understand patterns of violence in a specific way. In this report, we have highlighted the importance of gathering large amounts of detailed quantitative data on individual cases of missing persons and homicides. However, our task is incomplete without qualitative information, especially historical and political data that contextualize quantitative patterns of violence. Qualitative data ground our understandings of the causes of violence and, ideally, provide the theoretical basis for stratification and model selection.

The preliminary conclusions of this report have helped us refine questions for the national analysis. Together, these investigations will provide evidence regarding the magnitude, trends and patterns of violence in each department, contributing to the debate about violence in Colombia: is violence increasing or decreasing? Methodologically rigorous, theoretically grounded scientific analysis of violence can help enable honest dialogue to improve the human rights situation in Colombia.

About the Authors

Daniel Guzmán, B.S., is a statistical consultant for the Benetech Human Rights Program. He has contributed to project design and data analysis for Colombia, Guatemala and Sierra Leone. Mr. Guzmán served as the teaching assistant for Benetech's 2005 course on "Measuring Human Rights Violations" in Bogotá, Colombia. He received his B.S. in Statistics from the National University of Colombia.

Tamy Guberek, B.A., is the Latin America Field Coordinator for the Benetech Human Rights Program. Ms. Guberek supports partners with human rights information management systems and data analysis. She has also contributed to the descriptive statistical analysis for Benetech analysis of violence in Sierra Leone, and she manages the Benetech projects in Colombia and Guatemala. She received her B.A. in International Relations and Peace and Justice Studies from Tufts University.

Amelia Hoover, M.A., currently a Research Fellow at the Human Rights Program at the Benetech Initiative, is a Ph.D. candidate in political science

at Yale University. Her dissertation research focuses on human rights abuses during armed conflict, including covariation between lethal and non-lethal forms of violence and the effects of armed group command and control structures on patterns of violence.

Patrick Ball, Ph.D., is the Director of the Human Rights Program and Chief Scientist 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 Sri Lanka, Colombia, Burma, Liberia, Lebanon, Guatemala and in other countries around the world.

Our thanks to Beatriz Vejarano for matching and translation support, to all the organizations that contributed their data to this study, to Jeff Klingner for invaluable assistance with inter-rater reliability calculations, and to John Gilmore and Ann Harrison for supporting us during the final sprint to finish the report.

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 Omidyar Network, the Skoll Foundation, the John D. and Catherine T. MacArthur Foundation, and The Oak Foundation as well as project specific funding from EQUITAS and the Canadian Embassy.

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