

## Chapter 9

# Counting the Unknown Victims of Political Violence: The Work of the Human Rights Data Analysis Group

**Ann Harrison**

*Benetech Human Rights Data Analysis Group, USA*

### **ABSTRACT**

*The Benetech Human Rights Data Analysis Group (HRDAG) (<http://www.hrdag.org/>) analyzes the patterns and magnitude of large-scale human rights violations. Together with local partners, HRDAG collects and preserves human rights data and helps NGOs and other human rights organizations accurately interpret quantitative findings. HRDAG statisticians, programmers, and data analysts develop methodologies to determine how many of those killed and disappeared have never been accounted for - and who is most responsible. This account illustrates how HRDAG pioneered the calculation of scientifically sound statistics about political violence from multiple data sources including the testimony of witnesses who come forward to tell their stories. It describes methodologies that HRDAG analysts have developed to ensure that statistical human rights claims are transparently, demonstrably, and undeniably true.*

### **INTRODUCTION: DOES THE TRUTH MATTER?**

We see their pictures in news reports from conflict zones around the world. Frightened people, forced out of their homes, abducted, imprisoned, abused and sometimes killed by assailants who assume

they will never be held accountable for their actions. Stories of mass human rights violations stir our moral outrage. We feel compelled to act, to document the violence. How many atrocities have been committed? How many displaced into refugee camps? How many people have simply disappeared?

DOI: 10.4018/978-1-4666-1918-0.ch009

Data gathering is the cornerstone of human rights efforts. Information about the scope and patterns of violence is needed to support legal action, political sanctions, and humanitarian intervention. Yet this vital information is often clouded by biases or inaccurate data collection. These errors produce more than a distorted view of the facts. Inflated claims or false accounts that exclude key events can embolden critics and undermine well-intentioned efforts to prosecute perpetrators.

Consider the case of Kosovo. In March 1999, NATO allies launched air strikes against the Yugoslav army based on information from the U.S. State Department that 100,000 Kosovo Albanians were missing and feared dead. Members of the UN Security Council backed this decision arguing that the number of deaths justified action to halt further atrocities. After the bombing campaign and subsequent invasion of Kosovo by NATO, it was found that Yugoslav forces were responsible for about 10,000 deaths. Opponents of the action seized on the ten-fold inflation of possible deaths to dispute that mass violations occurred. Was there sufficient violence to justify war on humanitarian grounds? Do the numbers matter?

In an effort to defend the accuracy of human rights data – and account for mass atrocities witnessed by thousands or tens of thousands of people - truth commissions and other groups have turned to scientists to clarify history. Since 1991, members of the Human Rights Data Analysis Group (HRDAG) have developed relational data bases and statistical techniques to analyze not only how many victims and violations have been *documented* during a conflict – but also how many of the killed and disappeared have never been accounted for. HRDAG helps NGOs and human rights organizations accurately interpret quantitative findings that support historical clarification and the prosecution of perpetrators.

The HRDAG team currently includes two dozen statisticians, programmers, demographers and data analysts who passionately believe that

scientific arguments can help clarify the past and end impunity. HRDAG is part of the Human Rights Program at Benetech, a nonprofit technology organization based in Palo Alto, California. The HRDAG team is led by Dr. Patrick Ball, a statistician who has worked with dozens of human rights organizations to build databases that provide a quantitative understanding of human rights violations. HRDAG investigators focus on large-scale violations of civil and political rights. They have pioneered scientific methods for collecting and interpreting information about state-sanctioned violence that can be used to identify those responsible. The scientists of HRDAG have been called on to help clarify patterns of violence under difficult conditions in some of the world's most troubled nations. Over the past twenty years, HRDAG investigators have analyzed human rights data for official truth commissions in South Africa, Haiti, Guatemala, Perú, Ghana, Sierra Leone, Timor-Leste and Liberia. They provided critical information to the International Criminal Tribunal for the Former Yugoslavia that was used to prosecute former Serbian president Slobodan Milošević for crimes committed in Kosovo. (Ball, Asher, 2002) In 2010, HRDAG analysis of the Guatemalan National Police Archive was used to support the successful prosecution of two former Guatemalan police officers for the 1984 disappearance of a labor activist - one of the few examples of accountability in that troubled country. HRDAG has also analyzed information about human rights violations for the International Criminal Court and United Nations Field Missions in Timor-Leste, Guatemala, and the Democratic Republic of Congo, as well as numerous non-governmental human rights groups.

The methodologies that HRDAG has developed over the last twenty years are now being rediscovered and adapted by social scientists throughout the world. HRDAG analysts created these methodologies by carefully examining a wide variety of data in countries where political conflicts created chaotic descriptions of violence.

This chapter will illustrate how HRDAG investigators worked through the technical challenges of how to examine a wide range of information sources in which the obvious data is not a reliable narrator of the true pattern of violence. Each source of information must be carefully examined to ensure that its inherent biases do not skew an accurate historical clarification of events. As scientists, members of the HRDAG team remain non-partisan. They do not take sides in military or political disputes and they do not support the advocacy of any particular government or policy. HRDAG is not neutral, however. Members of the group are always in favor of human rights and support the promotion of the Universal Declaration of Human Rights and the International Covenant on Civil and Political Rights.

The scientific process seeks to remove researchers' prejudices from their findings. Using mathematics, relational databases and statistical analysis, the scientists of HRDAG can determine, with a high degree of certainty, which individuals are most responsible for crimes against civilians in a given conflict. But these efforts are based on the acknowledgement that observed events only offer a partial version of the truth. Scientists believe that statements can be made about occurrences that are true based on available evidence within a range of known probability. They base these findings on reproducible techniques that account for possible error and bias. HRDAG argues that scientific analysis of human rights violations can move arguments about history past political polarization and support policy initiatives that help heal troubled counties.

The diversity of conflicts in counties where HRDAG scientists work has required that they develop a range of solutions to carefully interrogate human rights data. HRDAG's investigations in Perú, Ghana, Sierra Leone, the Democratic Republic of Congo and other conflict zones has shown that by itself, large amounts of data shed only limited light on the truth of human rights conflicts. This chapter chronicles how HRDAG

has approached the analysis of complex of human rights information by coding narrative testimony and developing innovative surveys and statistical models to uncover obscured facts. It will illustrate how HRDAG creates complex relational databases to compare data sources and uses the statistical technique of multiple systems estimation to present accurate and defensible human rights data that can be used as the basis for policy decisions.

### **BACKGROUND: WHY IS HUMAN RIGHTS DATA SO DIFFICULT TO ANALYZE?**

HRDAG researchers rely on information from many different sources. They examine individual testimonies, legal depositions, administrative records from morgues and cemeteries, exhumation reports, operational records from prisons, career information on military and police officers, historical police archives, eyewitness interviews and official customs and immigration records. After locating these records, the HRDAG team designs statistical methods to analyze the information and find patterns in the violence.

An important part of HRDAG's mission is to point out to the human rights community why it's so difficult to accurately collect, structure and analyze this information. For instance, testimony by witnesses about assassinations may also include information about detentions or torture that victims experienced before they were killed. If detentions, torture and assassinations are all of interest to researchers, then all three types of violations must be represented in the researcher's database. Collecting information about only the "most important" violation will obscure instances of the other violations, creating significant biases in the analysis. Researchers must also consider both violations that were directly observed, and violations that were not witnessed but estimated based on the available information. (Ball, 2001)

## **Selection Bias**

In addition to under-counting incidents, many human rights data projects are vulnerable to selection bias. Every group collecting human rights data will have better access to some victims than others. This creates distortions in the statistical analysis of violations. The most frequent source of bias in human rights projects occurs because some victims or witnesses will trust or have access to the organization collecting data, while others will not. For instance, if victims in the southern part of a country trust investigators while those in the north do not, investigators will likely gather more testimonies in the south. This bias could result in a false finding that violence is primary based in the southern region. This kind of bias can also arise from limited access to unreachable or dangerous areas. If those analyzing human rights violations in a given country are basing their analysis on press accounts, and reporters avoid areas that they perceive as more dangerous, then the data about violence in those areas will be skewed.

Due to the nature of how human rights violations are witnessed, duplicate reporting of the same incident, victim or perpetrator also becomes a major problem for researchers. A human rights group may collect several accounts of one incident because multiple witnesses may report the same event. For example, the testimony describing an especially notorious assassination may be reported hundreds of times by different witnesses. This phenomenon was illustrated in witness testimony about the killing of former Liberian president Samuel K. Doe which HRDAG analyzed for the Liberian Truth and Reconciliation Commission (Cibelli, Hoover & Krüger, 2009). A good information system needs to identify which records describe the same events and participants in order not to double or triple count them.

## **Counting Errors**

Several sources recounting the same human rights violation may also provide slightly different versions of the same event. The interpretation of the actions or organizational affiliation of perpetrators, the location of a crime scene or the nature of a violation, as understood by a witness, may be different than the interpretation of these details provided by the organization that is collecting and analyzing the information. Different numbers of perpetrators may also have committed each violation. Consequently, no count of perpetrators from one organization can be combined with the number of perpetrators from other groups unless the perpetrators are combined in categories. Since any given violation may have been committed by one or more perpetrators who were not identified – perhaps by one, two, or many identified perpetrators - researchers must combine categories of perpetrators to analyze how actual violations were committed. Even when researchers consider these complexities, counts of perpetrators are still limited only to events that are directly observed.

Finally, it is difficult for researchers to accurately analyze data when witnesses offer contradictory views of the same event. HRDAG uses data analysis techniques to assess the underlying uncertainty about the incident in social memory. This allows analysts to recalculate statistics that describe a human rights violation using contradictory views of the same incident. If the results remain stable, investigators know that their data analysis can take into account possible variations on the same story.

The wide range of conflicts in which HRDAG analysts are asked to analyze human rights data compelled members of the research team to find scientific techniques that can address many different types of violence. Of the 19 countries where HRDAG has analyzed large-scale human rights violations, the most critical development of its methodologies took place in four troubled places, El Salvador, Guatemala, Kosovo and Timor-Leste.

## **EL SALVADOR 1991: WHO DID WHAT TO WHOM?**

In an effort to capture and accurately reflect the complex relationships between different people and events, Ball created a model for examining human rights data that he named “Who Did What To Whom” (WDWTW). Ball, who was originally trained as a sociologist, began analyzing human rights data in 1991. The initial focus of his study was the cauldron of state-sponsored terror carried out against the citizens of El Salvador. Ball looked for ways to analyze a large collection of human rights testimonies and determine who was most responsible for these crimes. It was an extraordinary challenge under difficult circumstances. In the early 1990’s, a 12-year civil war in El Salvador was coming to a close. The conflict between the military-led government and the Farabundo Martí National Liberation Front (FMLN), a coalition of left-wing militias, was sparked by the infamous assassination in March 1980 of Catholic Archbishop Óscar Romero. The Archbishop was shot while conducting mass, a month after publicly asking the U.S. Government to stop military aid to the Salvadoran government. At his funeral, bombers and snipers massacred forty-two mourners.

In the years that followed, tens of thousands of Salvadorans were killed, many by right-wing death squads dressed in civilian clothes and traveling in anonymous vehicles. The death squads conducted torture and disfigured bodies. They published death lists of intended targets, delivered coffins to future victims and sent them invitations to their own funerals. By the mid-1980’s state sponsored terror against civilians included indiscriminate bombing, mining and the harassment of medical personnel. Members of the security forces tortured and killed *campesinos* (rural peasants), shot at their houses and burned their crops. Refugees from the countryside poured into the capital city. Fighting continued through 1991 and peace accords were signed in late December of that year.

In an effort to understand the nature and scope of the conflict, the Salvadoran Human Rights Commission, founded in 1977 by Archbishop Romero, launched one of the earliest and most ambitious large-scale human rights information projects. Known in Spanish as the Comisión de Derechos Humanos de El Salvador no gubernamental (CDHES), the Commission collected over 9,000 testimonies from people who witnessed the violence. The testimonies helped begin to clarify what happened during the years of conflict. But individual stories were not enough to hold those who had ordered killings accountable for their actions. Ball, who had never before examined human rights data, was then living in El Salvador working for a peace organization. One day, he was approached by his human rights colleagues who asked him if it would be possible to organize their information in a database.

As a social scientist and database designer, Ball knew that he had to find some method to create units of analysis out of the unstructured narrative of the thousands of testimonies. Ball realized that the raw information generated from the testimonies would be complex. A single person could be identified as a witness, a victim and/or a perpetrator within a sequence of events. He knew that he needed a way to model the data that could accurately reconstruct which victims suffered which violations committed by which perpetrators. Ball worried that simplifying information about the violations could lead to distorted statistical results that could be attacked by opponents of the process. To capture and maintain these complex relationships between different people and events, Ball began creating a model for examining human rights data that he would later name, “Who Did What To Whom (WDWTM).”

Computer models are efficient, but they depend on carefully structured information. Before analyzing the patterns of the violations committed in El Salvador, Ball first created a structured, relational database of violations reported in the testimonies. Relational databases were first used



in the early 1970's on mainframe computers. By the early 1980s, PC users were building their own small relational databases that described complex relationships between different pieces of information. These types of databases are well-suited to counting and tabulating human rights violations and then creating statistics based on these counts. Turning to available commercial software, Ball began building his database of violations using commercial FoxPro database software.

Human rights groups started using databases in the mid-1980s. Some organizations built unstructured databases to keep track of individual cases. Other groups created text databases to gain easy access to documents. But these databases have limitations. A researcher using an unstructured text database to search for cases involving "electric shock," for example, will need to search for many different words or phrases that describe "electric shock." Many documents and case files are likely to emerge in this broad search and important references are missed. A structured database is different in that it describes violations using a controlled vocabulary. Any violation that is classified as an "electric shock" fits within the organization's definition of this kind of violation. A search in the database for "electric shock" can be absolutely specific and yield all of those cases - and only those cases - in which electric shocks are used. By framing the search differently, a researcher using a structured relational database can, for instance, list all those people who suffered electric shock torture, organized by names or by profession. Or the researcher can list all the police stations in which instances of this kind of torture have been alleged. If a researcher is searching for cases that occurred in a particular time, or at a particular place, a structured database can yield precisely these cases. Structured databases can generate simple or complex statistics that can be useful to human right groups examining thousands, or tens of thousands of individual testimonies. However, these statistics were subject to a key flaw that did not become obvious until the late 1990s.

In El Salvador, Ball used his WDWTW system to begin methodically building a structured, relational database using 200 cases from the 9,000 available testimonies to the Salvadoran Human Rights Commission. While Ball was working on this database, others in his human rights group began creating an index to the 9,000 testimonies using a table feature in WorldPerfect 5.1 DOS word processing software. Ball wrote a program to read the WordPerfect files and transform the tables into database records. But creating one database and an index was not enough to determine who was most responsible for twelve years of savage human rights violations in El Salvador. To address this question, Ball began creating another structured database that showed the career history of the 400 most senior military officers. This database held information describing when each officer graduated from military academy, what jobs they held, and which units they commanded. Much of the information was gleaned from newspaper accounts and declassified U.S. government documents.

While Ball was compiling these service records, another Los Angeles-based human rights organization named El Rescate, began creating their own database of Salvadoran military officers. El Rescate monitors human rights conditions in El Salvador and in 1985, began publishing reports on the country's human rights situation. The group developed a theoretical foundation for combating institutionalized impunity, established a legal basis for prosecuting past human rights violations and defined ways to utilize command responsibility to hold senior officers accountable for systematic human rights abuses (Howland, 2008).

In 1987, El Rescate began collaborating with El Salvador-based NGOs to petition international bodies such as the Inter-American Commission on Human Rights (IACHR) and the International Labour Organization (ILO). El Rescate also pushed for structural changes in the Salvadoran police and military and for the creation of entities that had the power to remove active security officers for human rights abuses - and identify specific

abuses and responsibility for them. This effort was successful due to a unique political moment created by the UN-brokered Salvadoran Peace Accords, signed in the closing minutes of 1991. The Accords mandated structural reforms, created a quasi-judicial Ad Hoc Commission to remove those implicated in human rights violations from military service, and formed a Truth Commission to compile an official public accounting of human rights abuses.

El Rescate shared their database of Salvadoran military officers with CDHES. Ball compared El Rescate's information with the officer data collected by the CDHES. The El Rescate database was one of first examples of "found data" that the HRDAG research team now routinely seeks in the countries where they work. This data comes from a wide variety of sources including cemetery or morgue records, border crossings, church groups or labor unions. Any group that collects data on human rights violations is a potential source of information.

After completing the two CDHES databases describing witness testimony and information about the Salvadoran military officers, Ball linked the data sets together. He then compared the dates in which individual officers lead specific military units with the dates of documented human rights violations. El Rescate also compared their officer data with a different database of violations acquired from the Salvadoran Catholic Church. Both sets of analysis identified 100 Salvadoran military officers who led units involved in the worst human rights violations. (Howland, 2008)

The results of this analysis were delivered to the Ad Hoc Commission which was empowered by the Peace Accords to force the named officers to resign. While the Salvadoran Peace Accords did not support legal trials, the names of the officers were reported in the press. Posters displaying their photographs and listing their crimes were displayed on the streets in the capital city of San Salvador. This information sparked a public uproar and the officers were forced into retirement. The

ultimate test of Ball's new WDWTW system took place when officers responsible for many of the killings and incidents of torture sued CDHES for defamation in Salvadoran court. A guilty verdict would have resulted in criminal penalties. Staff lawyers for the CDHES responded by bringing to court boxes of documents containing all 9,000 testimonies – plus the FoxPro software and the database which Ball offered to demonstrate for the judges. The officers withdrew their case.

### **GUATEMALA 1993 – 1999: USING MULTIPLE SYSTEMS ESTIMATION TO ANALYZE HUMAN RIGHTS DATA**

Inspired by the success of his human rights data analysis in El Salvador, Ball began to apply his Who-Did What-To-Whom (WDWTW) model to human rights information in other countries. Over the next six years, Ball was asked to analyze large-scale human rights violations in Ethiopia, South Africa, Haiti and Guatemala. Working in Washington, D.C, where he was a senior program associate for the Science and Human Rights Program at the American Association for the Advancement of Science (AAAS), Ball developed new methods for analyzing state-sanctioned violence. He was assisted by senior scientific colleagues including statistician Dr. Herb Spierer. This work helped Ball and his fellow researchers build on their experience in El Salvador and better understand the complexities of human rights data.

In 1992, a group of non-governmental organizations in Guatemala asked the scientific community to gather and analyze information on human rights violations. Ball advised the International Center for Human Rights Research in Guatemala (CIIDH) as they collected evidence of more than 43,000 human rights violations during the country's 36-years of conflict which ended in 1996. The descriptions of this violence that Ball gathered for the CIIDH were collected from press reports, NGOs and testimonies from

victims of the violence. From 1992 to 1998, Ball assembled this data into one large database and protected the information by encrypting it with PGP encryption software.

In 1996, peace accords brokered by the United Nations ended the armed conflict in Guatemala. Ball's data gathering project for the CIIDH provided a valuable source of information for the Commission for Historical Clarification (CEH) which was formed by the United Nations after the peace accords. The CEH sought to objectively clarify human rights violations and acts of violence in Guatemala by investigating a portion of the deaths and disappearances reported during the conflict. In 1998, the CEH asked the American Association for the Advancement of Science (AAAS) to answer the following question: How many people were killed in Guatemala from 1960 to 1996? Ball was asked to set up an information management system to analyze the large numbers of reported human rights violations and undertake the analytical study requested by the Commission.

When Ball began analyzing data about patterns of violence in Guatemala, he developed methods to overcome one of the basic complexities of counting human rights violations – removing duplicate reports of the same incident. A truthful accounting of human rights violations demands that every violation counts once and only once. Ball and other researchers began refining a process of de-duplication in which they could identify multiply reported incidents and assure themselves that each violation is counted only once. This de-duplication process, together with methods that control for bias in groups collecting human rights data, allowed Ball to use a statistical technique called multiple systems estimation (MSE). Dating back to at least the nineteenth century where it was first used to count fish, MSE was later extended to estimate both wildlife and human populations. MSE allows researchers to use separately collected but incomplete lists of a given population to estimate the total size of that group. It uses the identification of multiply-reported incidents – essentially

overlapping accounts - to estimate the number of events that are never reported.

When Ball was asked to analyze mass human rights violations in Guatemala, statistician Fritz Scheuren of the American Statistical Association, suggested that MSE could be used to project how many total violations took place, including those that were never documented. In a human rights calamity in which many thousands of people are killed or disappeared over many years, information about the uncounted victims is vital to producing a complete and accurate historical record and provided the evidence needed to document years of violations. This was especially necessary in Guatemala where violence had been ongoing for many years.

To answer the question of how many people died during the violence in Guatemala, Ball compared information from three available three databases of human rights information. Because state-supported violence in Guatemala was so extreme and had continued for so long, a number of different groups collected data on killings and disappearances. This allowed Ball to compare different collections of information and use MSE to determine how many victims had never been counted. The databases Ball used included the data set that he created for the (CIIDH). The second data set was generated by the CEH and the third was gathered by the REMHI (Recovery of Historical Memory) project which originated at the Human Rights Office of the Archbishop of Guatemala. Since Ball had three sources of data, he attempted to determine how many unduplicated killings were documented by the three projects and then calculate how many total killings occurred during the conflict. Ball completed a rudimentary de-duplication on the CIIDH data in August and September of 1998. He then conducted tests using the CEH, CIIDH and REMHI data in the fall of that year and completed the multiple systems estimation in January of 1999. (Chapman & Ball, 2001)

Duplicated reporting of deaths in more than one database is called “overlap.” To estimate the



total number of victims reported by the three databases, the overlap between databases must be estimated to reduce the contribution of each database by its overlap rate. If none of the victims in any database appear in any other database, then the total number of victims of killing and disappearance is equal to the sum of the number of victims in the three databases - which in this case was 54,643 victims. This is the upper limit to the number of observed victims. The lower limit can be found in the extreme possibility that the largest of the three databases, the CEH database, contains all the cases reported in the other REMHI and CIIDH databases. In this case, the total number of killings is simply the number of killings reported in the largest database which was 24,910. The true number of unique victims in the three databases must fall within these two limits, that is, between 24,910 and 54,643. The purpose of Ball's analyses to estimate the total number of documented killings was to narrow this range of possible victims.

Ball also knew that many killings that took place in Guatemala were not reported to any of the three projects collecting information. During the data gathering process, there were problems in the coverage of violations over time and location. None of the three projects provided good coverage of violations that took place from 1960 to 1977 and thus no estimation of this period was possible. The most important complication for the estimation of more recent violations was that the data gathering projects did not cover all the regions with the same intensity. In order to calculate the total number of killings with multiple systems estimation (MSE) Ball needed to generate estimates of the overlap between the three databases that collected data in the same areas.

Using MSE, Ball was able to calculate the probability of any given killing being documented by each database. It was then possible to estimate the number of killings excluded by two of the databases. With the same logic, Ball could then derive an estimation for the number of killings that were

not documented by any of the three projects. Using these techniques, Ball was able to calculate the total number of killings in Guatemala from 1979 to 1996 by region. Over all regions in Guatemala, Ball determined that there were approximately 47,706 documented killings. He was also able to calculate that there were approximately 84,468 killings that were not reported to the CEH, to the CIIDH, or to the REMHI project. Ball's final estimate, which was published as part of the CEH report in 1999, was that 132,174 killings took place in Guatemala between 1978 and 1996 with a standard error of 6,568. This number of deaths due to violence dwarfed the number of those killed in the publicized political conflicts in El Salvador, South Africa, Haiti and Perú. In the final analysis, the CEH also found that acts of genocide were committed against Guatemala's indigenous Mayan communities. MSE allowed Ball and his fellow analysts to account for those Guatemalans whose deaths had never been recorded (Ball et al., 2000, especially chapters 8-12).

### **KOSOVO 1999: USING MULTIPLE SYSTEMS ESTIMATION TO ANALYZE A VERY DIFFERENT CONFLICT**

After the first use of MSE to account for undocumented deaths in Guatemala, Ball then applied these techniques to clarify the history of a very different conflict in Kosovo. The turmoil began in 1989 when Serbian President Slobodan Milošević revoked Kosovo's autonomous status within the Republic of Serbia and sparked a violent confrontation between the Kosovar Albanians and the Yugoslav government. Allegations of widespread and systematic human rights violations were made against Serbian forces including deportations, destruction of homes and property, murder and rape. Ultimately NATO intervened to repel Serb forces from Kosovo. (Ball, Tabeau, Verwimp, 2007)

With support from members of the American Association for the Advancement of Science (AAAS), Ball gathered information in the field and wrote several reports on the conflict in Kosovo. First Ball, then Scheuren traveled to Albania in 1999, gathering data from border crossings and other sources. They then used this information to examine the claim by the Yugoslav government that the activities of both NATO and the Kosovo Liberation Army opposition force displaced the population and created a refugee crisis. During his analysis, Ball used MSE to examine the exodus of the refugees and compare this data with information about the scope of the conflict over time. (Ball, 1999)

The first report on migration in Kosovo was released in March 2000 by AAAS (Ball, 2000), and a second report on the killings was released a few months later with partners from the American Bar Association/Central European and Eurasian Law Initiative (AAAS & ABA CEELI, 2000). A third report was used in the prosecution of former Yugoslav president Slobodan Milošević at the International Criminal Tribunal for the Former Yugoslavia (ICTY) at The Hague.

The report given to the ICTY, "Policy or Panic? The Flight of Ethnic Albanians From Kosovo, March-May 1999," (Ball, Betts, Scheuren, Dudukovich & Asher, 2000) looked at possible reasons for the mass exodus of refugees from Kosovo. It examined the claim by Yugoslav authorities that military actions undertaken by NATO and the Kosovo Liberation Army (KLA) were responsible for the mass migration of Albanian refugees from Kosovo. The report found that killings and refugee flows occurred in a regular pattern characterized by three phases. In each phase, a high volume of killings and refugee flow was followed by a much lower level of killings and flow of refugees. The analysts used these findings to evaluate three possible explanations for the pattern. The first potential explanation is that military actions by the KLA forced Kosovars to leave their homes. The second possible answer was that NATO air at-

tacks created local conditions that led to Kosovars being killed and leaving their homes. The third potential reason for the pattern is that a systematic campaign by Yugoslav forces expelled Kosovar Albanians from their homes.

The study concluded that the pattern of both refugee flow and killings exhibit characteristics that are consistent with the presence of an external cause for these events. It found that the flow of refugees and the killings occurred in the same places at the same times. This implies that there was a common cause for both phenomena. The study found that two of the hypotheses that analysts considered to explain the pattern, the military activities of the KLA and bombardment by NATO forces, are inconsistent with the observed patterns of refugee flow and the killing of refugees. In their final analysis, the researchers concluded that the statistical evidence was consistent with the hypothesis that Yugoslav forces conducted a systematic campaign of killings and expulsions in Kosovo.

### **TIMOR-LESTE 2006: CREATING SURVEYS TO UNCOVER THE TRUTH**

In some countries where HRDAG analyzes human rights information, the available data is threatened or too incomplete for accurate historical clarification. In the small south Asian nation of Timor-Leste, HRDAG scientists took steps to preserve critical human rights information during civil unrest – and found new ways to combine found data and surveys to measure the impact of violence on civilian populations. Large-scale human rights violations in Timor-Leste stretch back to 1975 when the Indonesian government invaded the eastern part of the small, divided island. The resulting Indonesian occupation of Timor-Leste resulted in a series of abuses against the local resistance movement and the broader civilian population that lasted until the Timorese independence vote in August 1999. Disappearances,

torture, forced displacement and extra-judicial killings were documented during this period. The Timorese people also suffered a severe famine between 1978 and 1983. Qualitative historical accounts estimating the total death toll during the Indonesian occupation varied widely from a conservative 50,000 fatalities to more than 200,000 deaths. (Klingner & Silva, 2011, forthcoming) note how each individual data source, such as grave marker inscriptions, reflect only a fraction of the entire number of the dead. Such information is one kind of “found data” or records that were originally created or gathered for some purpose besides counting casualties.

In their paper, Silva and Klinger describe how the biases inherent in a single data source can be overcome by combining found data with surveys and other intentionally gathered data. This strategy allows analysts to more accurately estimate the magnitude and patterns of deaths. This methodology was tested when the Commission for Reception, Truth and Reconciliation in East Timor (CAVR) was established by the UN Transitional Authority in July 2001. The CAVR was tasked with investigating human rights abuses committed by all sides between April 1974 and December 1999, but it had little information at its disposal. HRDAG began advising the group on its information management processes and guided the development of the Commission’s statistical findings.

The partnership between the CAVR and HRDAG resulted in the creation of three data sets that integrated quantitative methods into CAVR’s broader truth seeking activities. These data sets included a database created by Ball, CAVR consultant Ken Ward and HRDAG demographer Romesh Silva. This database collected almost 8,000 narrative testimonies from people throughout the country who participated in the commission’s statement-taking process. Silva also conducted a census of all public graveyards in the country that encompassed approximately 319,000 gravestones. The third data set was comprised of a

retrospective mortality survey drawing on a probability sample of approximately 1,400 households throughout the thirteen districts of Timor-Leste (Silva & Ball, 2006).

To create these data sets, CAVR and HRDAG pioneered techniques for gathering human rights data. No previous truth commission had ever undertaken a survey to ask the population about deaths that occurred in the past. While historical demographers have used gravestone information to estimate deaths, this was the first time that a human rights project employed such methods. Each of the data sets independently produced valuable empirical results. Comparative analysis among the data sets corroborated the findings of each individual set of information. After matching deaths reported across all three sets of data, HRDAG analysts conducted multiple systems estimation to estimate the pattern and extent of conflict-related deaths and ultimately create a substantial body of documentary evidence to support the human rights findings of the CAVR. HRDAG also developed survey-based estimates of the extent and pattern of total conflict-related displacement between 1974 and 1999. The combined analysis of deaths and displacement complemented the Commission’s historical, legal and qualitative findings about famine and displacement in Timor-Leste. Finally, HRDAG developed an array of statistical data that profiled the form, pattern and structure of torture, ill-treatment, arbitrary detention and sexually-based violations which were reported to the CAVR.

These statistical and demographic findings, developed jointly by the CAVR and HRDAG, were presented in CAVR’s 2,500-page final report titled “Chega!” (Portuguese for ‘no more, that’s enough’). In February 2006, HRDAG co-published with CAVR a chapter of the report and accompanying mortality data (Silva & Ball, 2006). The report estimated that at least 102,800 (+/- 11,000) Timorese died as a result of human rights violations in Timor-Leste from a year prior to the Indonesian army’s invasion in 1975, to the

country's 1999 independence referendum that formally ended the occupation. HRDAG researchers estimated that that 18,600 people were murdered or disappeared, and approximately 84,200 citizens died due to hunger and illness in excess of what would be expected during peacetime.

HRDAG's analysis revealed the surprising truth that far more people died in Timor-Leste as a result of the famine which followed the conflict than in the fighting itself. (Silva & Ball, 2007) The actions that followed the release of the report were also unexpected. In July 2006, armed gangs broke into the offices of the CAVR in the capital city of Dili and stole their motorbikes. HRDAG scientists and other human rights observers were concerned that the mobs would soon return to loot the irreplaceable paper records used by the CAVR to compile the "Chega!" report. In case electronic records were destroyed or stolen, the CAVR and HRDAG created back-up copies of the CAVR's interview data and moved the information to secure servers. Other critical human rights data in UN offices was lost forever in the looting.

## **SOLUTIONS AND RECOMMENDATIONS**

Without combining data sources and estimation methods in Timor-Leste, a full understanding of deaths due to the conflict would have been incomplete. The number of people who died due to direct violence ultimately accounted for about 20% of deaths during the Indonesian occupation. By incorporating survey-based methods, HRDAG analysts were able to obtain more precise estimates of the full range of conflict-related deaths and gain new insight into deaths related to famine that were not as obvious.

After the damage caused by the riots in the Timor-Leste capital of Dili, HRDAG researchers released a statement calling for all current and future United Nations field missions deployed to Timor-Leste to avoid past mistakes and take

steps to secure and back up their human rights data outside of Timor-Leste. Despite the importance of information to their mission, many human rights organizations lack the resources to preserve their data securely. Much of their information is stored in a single hard disk, and often in unencrypted form. Critical documentation is often subject to viruses, computer theft, accidents, neglect and staff turnover. If files containing sensitive identifying information about victims and perpetrators are compromised, witnesses can be put at serious risk. HRDAG urges human rights groups to develop information management systems that store data electronically, in multiple copies and in multiple locations to prevent loss due to physical destruction. The analysts of HRDAG encourage their partners to encrypt data with Martus ([www.martus.org](http://www.martus.org)) or other software tools, and allow only authorized users to access the information.

HRDAG also cautions human rights organizations, and those who use their data, to be alert for bias in both information collection and analysis. Single sources of information offer only a limited view of what occurs during human rights violations and cannot be relied upon for unbiased analysis. Human rights analysts examining multiple sources of information should also take steps to correct for biases that result in the collection of only the "most important" violations. While this error can obscure the instances of other violations, even aggregated databases can contain significant bias. The who-did-what-to-whom (WDWTW) system that Ball developed in El Salvador beginning in 1991, addresses the creation of structured databases of human rights information. But directly reported information is always limited because only certain people chose to report the violence.

During their analysis of patterns of violence in Guatemala, HRDAG researchers also sought more efficient ways to remove duplicate accounts of violations. Much of the de-duplication of violations in Guatemala was done by hand, meaning that identifying overlaps required individual human expertise applied to each and every



potential match. But this process poses a number of problems. The number of people, the time and concentration this requires makes it impossible to analyze some large collections of data. In addition, criteria for hand-matching are necessarily subjective and inconsistently applied. It is impossible to eliminate the different judgments people use in evaluating data, even over the course of a single person's work. Most importantly, hand-matching is not auditable or repeatable. Those reviewing the analysis cannot do the matching themselves to validate the results, and researchers cannot review the prior matches when incorporating new or updated data. For researchers on the HRDAG team, improving the de-duplication process has been the key to expanding their analysis of human rights data. Over the years, they have developed a usable, reproducible large-scale system for automatically matching duplicated records of the same violation, victim or perpetrator.

While the scientists of HRDAG have focused on developing data analysis techniques that stand up to scrutiny, their findings and methods have been questioned in legal cases. One legal body that struggled with HRDAG's data is the International Criminal Tribunal for the Former Yugoslavia (ICTY) where Ball served as an expert witness for the prosecution in the trials of Slobodan Milosevic (ICTY IT-02-54) and Milan Milutinovic et al (ICTY IT-02-57) (Ball et al. 2002a, 2002b). In a report published in 2000 by the American Association for the Advancement of Science (AAAS), Ball analyzed four incomplete data sources on killings and forced migration during the conflict in Kosovo to estimate the pattern of deaths from March to May 1999. Ball's analysis identified 4,400 deaths from March to May 1999. He used this data to estimate that approximately 10,000 Kosovar Albanians had been killed during the March-May period.

After the 2000 AAAS report, Ball conducted another study for the ICTY (Ball et al. 2002a, 2002b) where he looked at other sources, including border-crossing records, to estimate patterns

of migration during the March-May period. He determined that the estimated patterns of deaths and migration were consistent with the hypothesis that Serbian forces had carried out these crimes. Milosevic died during his trial and the defendants were found guilty of most of the charges against them, but the ICTY were not convinced by Ball's statistical evidence. In her paper entitled, *Learning the Hard Way at the ICTY: Statistical Evidence of Human Rights Violations in an Adversarial Information Environment*, Amelia Hoover Green, notes that statisticians need to be sensitive to the overwhelming amount of information that judges must contend with during trials (Hoover, 2010). As the judges in these cases tried to determine the defendant's responsibility for large scale killings and migration of Kosovo Albanians during the March to June time period, Hoover suggests that statisticians may have made their findings clearer if they presented background data vital to the judge's understanding, conducted redundant and exhaustive hypothesis testing, focused on causation rather than description when possible, privileged simple, familiar or traditional modes of statistical investigation, pursued total transparency in data analysis process and trained to present their work in court.

HRDAG scientists have carefully considered Hoover's analysis of their work in these cases, and taken other steps to make the findings accessible and credible. In order to develop findings that stand up to scrutiny, all the software tools that HRDAG uses to analyze human rights data are open for review and critique. HRDAG employs open source software, such as the R statistical software, and also creates free software that uses open data standards. Over time, HRDAG has worked to publish the data used in its analysis, and the group is now working on publishing more of the software it uses to conduct its calculations. HRDAG researchers believe that transparency helps generate trust in the scientific process. All of HRDAG's findings are publicly available, and whenever possible, the results are presented



in reviewed scholarly journals. This is the way HRDAG believes that the science of human rights data analysis should be conducted.

## **FUTURE RESEARCH DIRECTIONS: ANALYZING A LARGE POLICE ARCHIVE**

HRDAG is currently involved in an unusual data project analyzing a sample of the estimated 31.7 million papers, books, photographs and floppy disks in the Guatemalan National Police Archive (Guzman, Guberek, Shapiro, Zador, 2009). The archive was discovered by chance in a military munitions storehouse near Guatemala City in 2005. The Guatemalan National Police were disbanded after the country's 1996 Peace Accords which ended 36 years of internal armed conflict. In March 2006, a coalition of Guatemalan NGO and government groups invited Ball to evaluate the documents in the collection. After inspecting the archive, Ball recommended that analysts collect a multi-stage random sample of the documents and secure the sample data with the *Martus* software for later analysis (Price, Guberek, Zador, Shapiro, 2009). With advisors from the American Statistical Association and Westat, HRDAG statisticians Romesh Silva and Daniel Guzmán designed a sampling strategy that is now being used to create a clearer picture of archive contents. Together with other HRDAG statisticians, Guzmán has been analyzing a sample of the documents in the archive generated from 1960 to 1996 during Guatemala's years of conflict.

Expert testimony by Guzmán provided key evidence in the October 2010 conviction of two former police officers who were found guilty of the 1984 forced disappearance of Guatemalan student and union leader Edgar Fernando García. The Recovery of Historical Memory (REMHI) Project led by the Catholic Church estimated that approximately 40,000 Guatemalans disappeared during the years of violence. Documents

discovered in the archive may help determine what happened to García who vanished in 1984 while in police custody. During the García trial, Guatemalan and international experts testified about National Police policies to "control and clean" suspected insurgents and their blurring of "insurgents" and political opponents. The analysis that Guzmán presented in his testimony was based on three years of coding key variables from random probabilistic samples of documents drawn from the archive. Since the contents of the archive were too large to directly observe, HRDAG statisticians developed a plan to create a physical map of the archive and sample randomly from different groupings of documents inside the building. Using this methodology, analysts were able to extrapolate from what was known about the sampled documents and make estimates about the contents of all the documents in the collection.

The García trial is the first case in the Guatemalan judicial system based primarily on archive documents. The guilty verdicts in this case have paved the way for judges to trust these records as evidence in future trials and accept statistical evidence which is rarely used in Guatemalan legal cases. Through his analysis, Guzmán was able to calculate the percentage of documents about García known by different police units and support the prosecution's argument that high-level officers were involved in his disappearance. This analysis continues to generate quantitative results that can answer further questions about command responsibility and human rights abuses. Analysis of police documents by Guzmán provided critical information used to support the June 2011 arrest of the former chief of the Guatemalan National Police who is accused of complicity in García's disappearance. The arrest of Hector Bol de la Cruz, 70, at his home in Jutiapa southwest of Guatemala City on June 9, is the latest step in an ongoing investigation of former police officials accused of disappearing and murdering García and other political activists.

By determining which documents were known by police command structures, HRDAG analysts can also make findings about relationships among security forces and communications between the army and police. This evidence is critical because historical data has shown that the Guatemalan army was the force most involved in human rights violations against civilians. The guilty verdict against the former National Police officers provides a critical legal precedent needed to prosecute others for forced disappearance. Judges in the García trial determined that Guzmán's analysis of archive records was relevant in building the case and removing reasonable doubt about the authenticity and reliability of the documents. Prosecutors have stated that HRDAG's statistical estimations regarding the content of archive documents will play a key role in future investigations. In particular, judges highlighted the role of this data in corroborating information about the disproportionate urban focus of the National Police, their practice of disappearing opponents, and the estimated number of documents in the archive related to disappearances. The judges in the case have announced that prosecutors will continue to investigate responsibility for García's disappearance further up the National Police chain of command.

## **CONCLUSION**

The scientists of HRDAG believe that the legitimacy of the human rights community comes from presenting uncomfortable facts to those in power. But demands for accountability after violent conflicts are often denied or dismissed. In the absence of firm evidence, these debates are paralyzed by competing claims of historical accuracy. In order to prosecute perpetrators of human rights violations, the courts rightly demand evidence. In crimes against humanity, researchers must show that these actions follow a pattern. Carefully crafted human rights arguments offer

the greatest opportunities to halt the cycle of violence and create lasting social change during difficult-to-foresee periods of political transition. Unless the human rights community is ready with unimpeachable information about past abuses, opportunities for official acknowledgment, accountability or reform are lost.

HRDAG data analysts have worked through an arc of technical issues to analyze human rights data which a new generation of statisticians are now reinventing and rediscovering. The development of these structures is important to chronicle because of how diverse and heterogeneous the problems have been and the scope of their solutions. In their quest to accurately measure mass human rights violations, HRDAG has discovered that human rights violations or events are complex and rarely straightforward. Nevertheless, NGOs and government groups have used available technology (even spreadsheets) to tell stories with data. The researchers of HRDAG keep reminding the human rights community and their supporters that data can be an unreliable narrator, a trickster. The data may sometimes reveal the truth. But if researchers rely on only a single source, such as press accounts or testimonies given to a single organization, they cannot be exactly sure when data is consistent with reality, and when it provides a reflection distorted by selection bias. As in the post-modern literary tradition, seemingly obvious information must be carefully interrogated. Investigators can examine data in complex ways, but larger quantities of information do not necessarily shed greater light on reality.

Careful analysis of human rights data can provide clues to the fates of people who disappear during conflicts and provide an opportunity to understand the role of political policies in disappearances, kidnappings and assassinations. This analysis can offer a view of the perpetrators' perspective on human rights violations and help researchers understand how a national leader, or their security forces, can turn against their own citizens. Researchers can defend these findings

by showing that their analysis is calculated from a consistent application of a scientific classification criteria applied to raw data. These scientifically defensible techniques can be used by prosecutors, as they were in the García case in Guatemala, to determine chain of command and responsibility for crimes of policy.

The scientific methods developed by HRDAG analysts over the last twenty years have set standards of scientific rigor that can help overcome political arguments about the content of human rights records. Together with its local partners, who often take great risks to do this work, HRDAG has established scientific paradigms for examining large collections of human rights data. These techniques have enabled scientists to transform emotional debates about politics, policies and history into evidence-based discussions about science and methodology. The researchers of HRDAG have shown that careful, accurate data analysis can help fight impunity, bring perpetrators to justice and strengthen the rule of law.

## REFERENCES

American Association for the Advancement of Science. (2000). *Political killings in Kosova/Kosovo, March-June 1999: A cooperative report by the Central and East European Law Initiative at the American Bar Association and the Science and Human Rights Program of the American Association for the Advancement of Science*. Washington, DC: Author.

Ball, P. (1999). *Policy or panic? The flight of ethnic Albanians from Kosovo*. Washington, DC: American Association for the Advancement of Science.

Ball, P. (2000). *Refugee flow patterns in Kosovo, March-May 1999*. American Association for the Advancement of Science, *Science and Human Rights Program*. Washington, DC: American Association for the Advancement of Science.

Ball, P. (2001). Making the case: The role of statistics in human rights reporting. *Statistical Journal of the United Nations Economic Commission for Europe*, 18(2-3), 163–174.

Ball, P., & Asher, J. (2002). Statistics and Slobodan. *Chance*, 15(4), 17–25.

Ball, P., Tabeau, E., & Verwimp, P. (2007). *The Bosnian book of the dead: Assessment of the database (Full Report)*. *Households in Conflict Network Research Design Note 5*. Brighton, UK: Institute of Development Studies, University of Sussex.

Chapman, A. R., & Ball, P. (2001). The truth of truth commissions: Comparative lessons from Haiti, South Africa, and Guatemala. *Human Rights Quarterly*, 23(4), 1–42. doi:10.1353/hrq.2001.0005

Cibelli, K., Hoover, A., & Krüger, J. (2009). *Descriptive statistics from statements to the Liberian truth and reconciliation commission. A Report by the Human Rights Data Analysis Group at Benetech and Annex to the Final Report of the Truth and Reconciliation Commission of Liberia*. Palo Alto, CA: Benetech.

Guzman, D. R., Guberek, T., Shapiro, G. M., & Zador, P. (2009). *Studying millions of rescued documents: Sampling plan at the Guatemalan National Police Archive (GNPA)*. *JSM Proceedings, Survey Research Methods Section*. Alexandria, VA: American Statistical Association.

Hoover, A. G. (2010). Learning the hard way at the ICTY: Statistical evidence of human rights violations in an adversarial information environment. In Smeulers, A. (Ed.), *Collective violence and international criminal justice: An interdisciplinary approach*. Antwerp, Belgium: Intersentia.

Howland, T. (2008). How El Rescate, a small nongovernmental organization, contributed to the transformation of the human rights situation in El Salvador. *Human Rights Quarterly*, 30(3), 703–705. doi:10.1353/hrq.0.0019

Klinger, J., & Silva, R. (2010). *Combining found data and surveys to measure casualties. Human rights data analysis group*. Palo Alto, CA: Benetech.

Price, M. E., Guberek, T., Guzman, D. R., Zador, P., & Shapiro, G. M. (2009). A statistical analysis of the Guatemalan National Police archive: Searching for documentation of human rights abuses. In Proceedings, J. S. M. (Ed.), *Survey Research Methods Section*. Alexandria, VA: American Statistical Association.

Shapiro, G. M., Guzman, D. R., Zador, P., Guberek, T., Price, M. E., & Lum, K. (2009). Weighting for the Guatemalan National Police Archive Sample: Unusual Challenges and Problems. In Proceedings, J. S. M. (Ed.), *Survey Research Methods Section*. Alexandria, VA: American Statistical Association.

Silva, R., & Ball, P. (2005). On the use of sample surveys and multiple systems estimations in assessing large-scale human rights violations: Recent experiences from Timor-Leste. *Proceedings of the Social Statistics Section – Joint Statistical Meetings*. Alexandria, VA: American Statistical Association.

Silva, R., & Ball, P. (2006). *The demography of large-scale human rights atrocities: Integrating demographic and statistical analysis into post-conflict historical clarification in Timor-Leste*. Paper presented at the 2006 meetings of the Population Association of America. Silver Spring, MD: Population Association of America.

Silva, R., & Ball, P. (2006). *The profile of human rights violations in Timor-Leste, 1974-1999. A report by the Benetech Human Rights Data Analysis Group to the Commission on Reception, Truth and Reconciliation*. Palo Alto, CA: Benetech.

Silva, R., & Ball, P. (2007). The demography of conflict-related mortality in Timor-Leste (1974-1999): Empirical quantitative measurement of civilian killings, disappearances & famine-related deaths. In Asher, J., Banks, D., & Scheuren, F. (Eds.), *Statistical methods for human rights*. New York, NY: Springer. doi:10.1007/978-0-387-72837-7\_6

## **ADDITIONAL READING**

Ball, P. (2005). On the quantification of horror: Field notes on statistical analysis of human rights violations. In Davenport, C., Johnston, H., & Mueller, C. (Eds.), *Repression and mobilization*. Minneapolis, MN: University of Minnesota.

Davenport, C., & Ball, P. (2002). Views to a kill: Exploring the implications of source selection in the case of Guatemalan state terror, 1977-1996. *The Journal of Conflict Resolution*, 46(3), 427–450. doi:10.1177/0022002702046003005

Gohdes, A. (2010). *Different convenience samples, different stories: The case of Sierra Leone*. Palo Alto, CA: Benetech.

Guberek, T., Guzmán, D., Price, M., Lum, K., & Ball, P. (2010). *To count the uncounted: An estimation of lethal violence in Casanare. A Report by the Benetech Human Rights Program*. Palo Alto, CA: Benetech.

Guzmán, D., Guberek, T., Hoover, A., & Ball, P. (2007). *Missing people in Casanare (Los Desaparecidos de Casanare)*. Palo Alto, CA: Benetech.



Lum, K., Price, M., Guberek, T., & Ball, P. (2010). *Measuring elusive populations with Bayesian model averaging for multiple systems estimation: A case study on lethal violations in Casanare, 1998-2007*. Palo Alto, CA: Benetech.

Price, M., & Guzmán, D. (2010). Comments to the article. In D. Mejia & M. J. Uribe (Eds.), *Is violence against union members in Colombia systematic and targeted?* Bogota, Colombia: Universidad de los Andes.

Silva, R., Klingner, J., & Weikart, S. (2010). Measuring lethal counterinsurgency violence in Amritsar District, India using a referral-based sampling technique. In Proceedings, J. S. M. (Ed.), *Social Statistics Section*. Alexandria, VA: American Statistical Association.

Silva, R., Klingner, J., & Weikart, S. (2010). *State coordinated violence in Chad under Hissene Habre: A statistical analysis of reported prison mortality in Chad's DDS prisons and command responsibility of Hissene Habre, 1982-1990. A Report by Benetech's Human Rights Data Analysis Group to Human Rights Watch and the Chadian Association of Victims of Political Repression and Crimes*. Palo Alto, CA: Benetech.

Silva, R., Marwaha, J., & Klingner, J. (2009). *Violent deaths and enforced disappearances during the counterinsurgency in Punjab, India: A preliminary quantitative analysis. A Joint Report by Benetech's Human Rights Data Analysis Group & Ensaaf, Inc*. Palo Alto, CA: Benetech.

## KEY WORDS AND DEFINITIONS

**Human Rights Data Analysis Group (HRDAG):** A group of two-dozen statisticians, programmers, demographers and data analysts. HRDAG is part of the Human Rights Program at Benetech, a nonprofit technology organization based in Palo Alto, California.

**Multiple Systems Estimation (MSE):** A statistical method that allows researchers to use separately collected but incomplete lists of a given population to estimate the total size of that group. MSE uses the identification of multiply-reported incidents – essentially overlapping accounts - to estimate the number of events that are never reported.

**Who Did What To Whom (WDWTM):** A model for examining human rights data created by HRDAG founder Dr. Patrick Ball.