ASKING AND ANSWERING HARD QUESTIONS: TECHNOLOGY IN THE SERVICE OF HUMAN RIGHTS

BY PATRICK BALL AND ANN HARRISON[1]

Sophisticated technology is increasingly making it possible to transform human rights arguments from political polemic to scientific debate.

How many people are imprisoned in China? Is the number increasing or decreasing?

If the names of all Chinese prisoners were made public, monitoring patterns of detention would be easy. But because this information is classified, campaigns to change state policies governing the detention of political dissidents can only be based on anecdotes.

Human rights activists now have stronger tools at their disposal. Statistical techniques developed in the last 15 years have helped turn human rights narratives into scientific estimates of the total magnitude and patterns of human rights abuses. These techniques have been used successfully in eight truth commissions and in several trials at the International Criminal Tribunal for the Former Yugoslavia.

Statistical evidence is especially good at considering whether human rights crimes could have been the result of deliberate policy or if they could have occurred through the actions of a few rogue soldiers.

For example, during the conflict between NATO and Yugoslavia in March–June 1999, there were some who argued that hundreds of thousands of Kosovar Albanian refugees left Kosovo because they were afraid of the NATO bombings. Through a statistical analysis described later in this article, we showed that village by village, day by day, refugee movements usually preceded the NATO attacks by days or weeks. Consequently, we concluded that NATO's airstrikes could not have been the cause of the refugees' decisions to leave their homes.

In other examples, we have established the relative proportions of killings committed by the state and by guerrillas (Perú), and the relative impact of killings on indigenous and non-indigenous communities, which helped to establish that the Guatemalan Army committed acts of genocide against the Mayan people.

Most recently, we have been able to estimate the different kinds of mortality that affected East Timorese civilians during the Indonesian occupation (1975–1999)—violence versus disease and hunger.[2] In each example, our work helped to clarify the responsibility not of individual perpetrators, but of entire policies and institutions.

Thinking about China

Let's go back to the question of how many people are imprisoned in China.

Human rights activists must first ask how many people are now *known* to be in prison. For instance, a human rights group might maintain a database of several thousand prisoners in China, yet it is certain that there are hundreds of thousands, perhaps millions more people in custody.

It is impossible to draw any statistical conclusions based solely on a hypothetical database. In order to understand the dynamics of detention—such as increases or decreases in the total population of prisoners, increasing or decreasing sentence lengths, or regional shifts in the number of prisoners held—organizations must find a method that enables them to estimate the *total* population of prisoners in each year, in each region and with each sentence length. This total must include both the prisoners whose fates are known as well as an estimate of those who have never been documented.

How would the human rights community do this? The standard statistical technique would be to draw a random sample of prisoners in China. This is clearly impossible. It might, however, be possible to "reverse engineer" the number of prisoners by using multiple independent, direct and indirect sources of information about them.

For instance, are there any publicly accessible court records that could be used to create lists of newly condemned prisoners? Are trials and sentences reported in newspapers? Careful reviews of publicly accessible materials have been successful mechanisms for tracing officially sanctioned human rights crimes in some countries.

Are there family support groups that could report who is in the prison, when they went in, on what charges and for how long? Could former prisoners be interviewed and asked the names of all their cellmates or fellow prisoners? Are former prisoners aware of executions that may have taken place inside prisons where they were held? Do some prisons have cemeteries with marked graves?

Each of these data sources provides a reference to specific individuals. This is useful for a technique known as *multiple systems estimation* that allows for a direct estimate of the number of prisoners held in each prison. If we can determine which individuals are on each of the lists, and which individuals are on several of the lists, we can estimate how many individuals are on *none* of the lists. This enables us to estimate the total population (of executions, of prisoners) so that we can analyze how that population may be changing.

There are also indirect methods for making estimations of the totals. These techniques do not necessarily permit us to estimate the number of people in a prison, but they can provide a reference for the trend of the population to increase or decrease over time. The indirect indicators may also help us see broad comparative variations in the scale of different prisons.

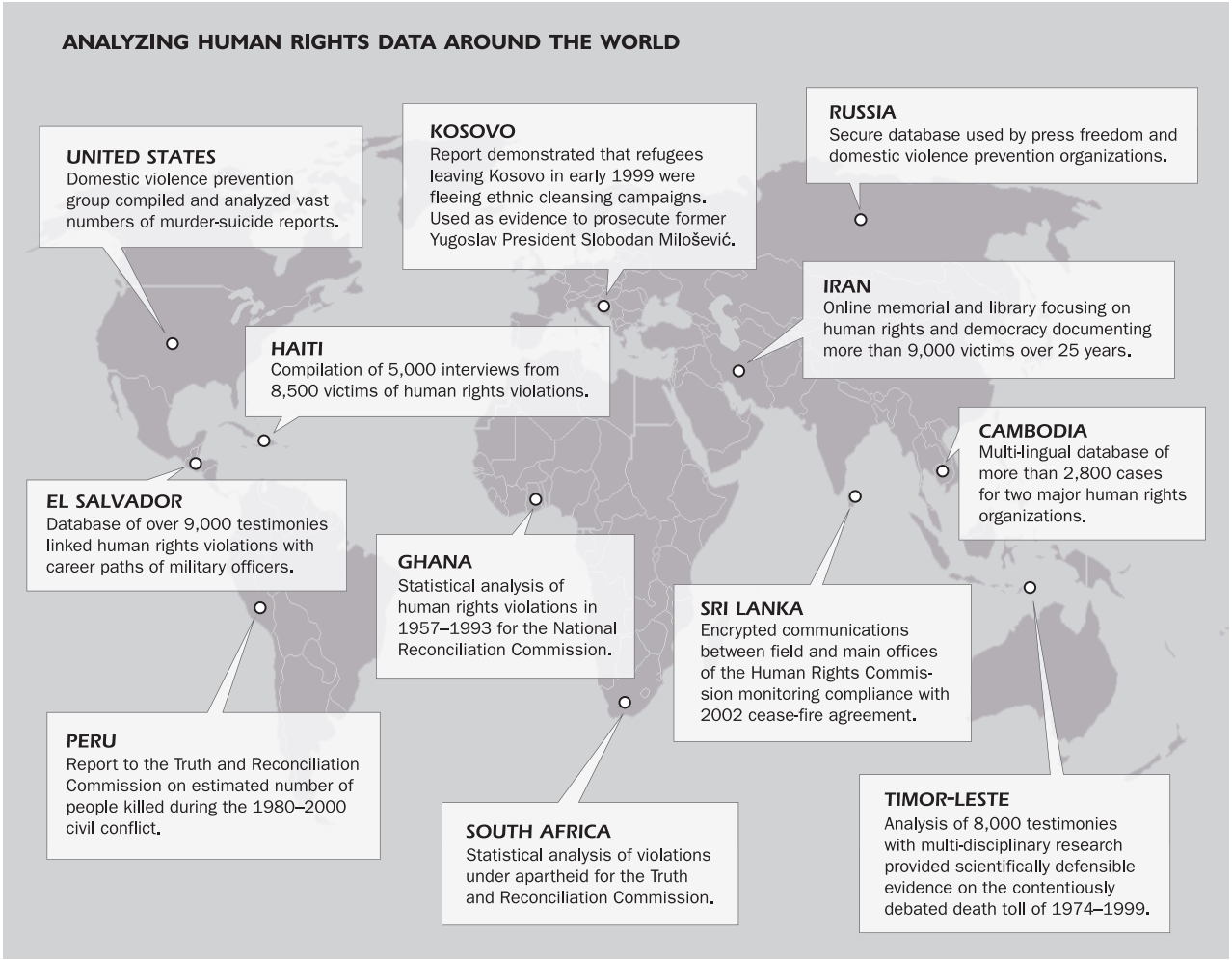
Do the vendors who supply food to prisons keep records that would allow us to observe increases and decreases in food purchases that would provide clues about the prison population? If one prison is purchasing twice as much food as another prison, for example, this would show that either the prisoners in the first prison are eating twice as much, or that there are twice as many prisoners in the first prison.

If we are looking at prisons where forced labor takes place, for example, we might be able to monitor the output of their products, such as the number of garments or circuit boards. Shipping companies that provide raw materials to the prison or transport finished products from the prison might have records that are accessible.

Fitting the data together

Creating quantitative analysis from many indirect partial sources is a bit like putting together a puzzle. Each piece of data—family reports of prisoners or daily food deliveries to the prison—provides another piece of the story. It is the analyst's job to combine all the sources and find common patterns that allow them to reject explanations or hypotheses that are inconsistent with the data.

Observational data and statistical techniques of this kind do not permit us to make affirmative proofs. We cannot say that statistical analysis proves that some argument is true. We can say that the data are consistent with a given hypothesis, but there may be other hypotheses that are also consistent with the data. A complete human rights argument must therefore incorporate several forms of evidence such as legal, historical, quali-



The concepts and software developed by Benetech Initiative's Human Rights Data Analysis Group (HRDAG)has been used by individuals and groups worldwide. Graphic by Shirley Hao

tative or documentary materials in order to make a conclusive finding.

This method of statistical analysis excludes hypotheses that cannot be true, leaving a much smaller number that could possibly be true. It is the analyst's job to combine the statistical argument with other forms of evidence to draw a conclusion.

The construction of an argument by fitting pieces together occurs at several levels. Suppose, for example, that the estimates we make from family stories and former prisoners' accounts suggest that the prison population increased for three years and then decreased thereafter. If we find a similar pattern of increase or decrease in the food supplies over time, the two independent findings greatly strengthen each other. If we then also find official policy declarations indicating a crackdown in the years preceding the prison population increase, we have another independent indicator strengthening the case of prison population dynamics.

Our objective is to build an argument that uses several different kinds of quantitative information that make the same point. Then we add other kinds of evidence—including historical, documentary, policy or journalistic—that lead to the same conclusion. The point is that no matter what approach a reader brings to the argument, the conclusion is the same.

What is the first step?

During the 1999 conflict in Kosovo between NATO and the government of Yugoslavia, hundreds of thousands of refugees fled Kosovo for safety in Albania and Macedonia. Why did they leave? To answer this question, the authors of this article and their colleagues used 11 independent data sources, including Albanian government border records, United Nations estimates for total numbers of refugees, random sample surveys conducted in refugee camps, narrative victim accounts and Yugoslav government Web sites that denounced NATO airstrikes.[3]

A complete human rights argument must incorporate several forms of evidence in order to make a conclusive finding.

We used nine distinct statistical techniques, three programming languages and tens of thousands of lines of software code to conduct the analysis. None of this would have been possible if the original material had not been collected, but we want to emphasize that very little of this information was collected specifically for this project. Instead, we found data collected by administrative agencies, other NGOs and the Yugoslav government, and transformed it into structures we could use in our statistical argument.

The most important step in any large-scale human rights data project is to begin preserving individual points of information. Most human rights data projects begin with individual victim accounts. Human rights groups are rapidly learning about other forms of information that contribute to advocacy based on scientific statistical techniques—but the first step is always to accumulate as much data from as many diverse sources as possible. The Benetech Initiative's Human Rights Program has published a software tool that helps human rights investigators manage raw information. This software, called Martus, secures and preserves valuable information in a way that makes it difficult for anyone to destroy.[4]

Securing the evidence

The job of a human rights investigator is to gather all data that could possibly be relevant and store it in a way that is accessible to colleagues, secure from the perpetrators and difficult to destroy.

In an effort to transform a human rights argument from political polemic to a scientific debate, human rights analysts can use tools adapted from computer science, mathematics, statistics and demography to make estimates of the total populations of prisoners, executions or victims of torture—if we have adequate data from qualified sources.

We must ask the questions that help us see how violence is used as a political tool. When we make arguments about governments that use systematic violence as a policy tool, we look for patterns that show a relationship between political changes and the practice of violence.

We must prepare for the day when historical truth will guide China's leaders to respect human rights. But when the moment comes to transform society, reformers must have the raw data for an evidence-based reflection of past violence.

Only with a thoughtful history can we fully respect the victims of political violence. Mythologies about idealized victims are often groundwork for the next round of persecution. In contrast, a rich, accurate history, with all its complexity, will help future leaders to reject the temptation to repeat the atrocities of the past.

NOTES

1. The use of the first-person plural "we" in this article refers to the entire staff of the Benetech Human Rights Program. Please see http://hrdag.org/about/people.shtml.
2. The Kosovo example is described in more detail later in the article. Our Perú report is available online http://shr.aaas.org/peru/aaas_peru_5.pdf; the Guatemalan analysis is available online at http://shr.aaas.org/mtc/chap11.html; and our report on Timor-Leste is available at http://hrdag.org/resources/timor_chapter_graphs/timor_chapter_page_01.shtml.
3. See Patrick Ball and Jana Asher, "Statistics and Slobodan," Chance, Vol 15, No. 4, 17-24 (2002) for a non-technical discussion of the methods used. The complete report submitted to the ICTY is here: http://shr.aaas.org/kosovo/icty_report.pdf (as of 18 May 2006).
4. For more information about the Martus platform, see http://www.martus.org.