



My brilliant career

Romesh Silva

Where can a mathematics education lead you? In this new series, mathematics and statistics graduates from Australian universities write about their careers, proving that the world is their oyster. In the first instalment of this series Romesh Silva describes a day in the life of a human rights statistician. Romesh Silva has worked on human rights data analysis projects in Timor-Leste, Sri Lanka, Chad and Palestine. He is a graduate of the University of NSW and Columbia University.

Quantitative data analysis of large-scale human rights violations: An example of applied statistics at the grassroots

It is 5am on a balmy Sunday morning in Dili. The roads are still deserted and the sun is only just starting to peep over the towering mountains which form the eastern perimeter of the town. We wearily bundle into the jeep and commence our long journey along the winding mountain roads of Timor-Leste to the tiny village of Mauxiga.

Only 8 hours ago did we receive a message from a colleague that the village was accessible. There are no telephones over in Mauxiga, and one of our colleagues had spent his entire Saturday traversing the mountains of Mauxiga looking for Abilio dos Santos. Abilio, the village secretary of Mauxiga has sent us a set of detailed lists of detentions, displacements and conflict-related deaths experienced by the people of Mauxiga. He would like us to include this data in the statistical analysis we are developing for the UN-mandated truth commission. In my preliminary descriptive analysis of the lists, I have identified several ambiguities and some key pieces of data which are missing. Hence, our efforts to locate Abilio, the principal author of the lists, to assist us with our analysis and clarify the ambiguities in the data.

Since September 2001, I have been working with a team of statisticians and database programmers in developing scientifically defensible measurements of the extent and pattern of large-scale human rights violations. Our team, the Human Rights Data Analysis Group (HRDAG) designs and builds information management solutions and conducts statistical analysis in partnership with truth commissions, NGO projects and UN missions¹. In June 2003, we were invited by the Commission for Reception, Truth and Reconciliation (CAVR, by its Portuguese acronym) to assist in integrating statistical science into the Commission's multidisciplinary research processes. CAVR was established in July 2001 by the United Nation's Transitional Authority in East Timor, through collaboration with the National Council and Cabinet after public consultation. CAVR is an independent statutory authority, mandated to inquire into human rights abuses committed by all sides between April 1974 and December 1999, in addition to facilitating reconciliation and justice for less serious offenses. CAVR's mandate period covers the internal civil war between Timorese political parties,

¹HRDAG was initially incubated at the American Association for the Advancement of Science. The Group is now based at the Benetech Initiative in Silicon Valley. See <http://www.hrdag.org> for details.



Reviewing some raw data and examining some time-series graphs in Mauxiga - Olga da Silva, Abilio dos Santos, Romesh Silva (seated) and Antonio Pires and Xavier da Amaral (standing).

the 24-year occupation of Timor-Leste by the Indonesian military, a famine between 1975 and 1982 and the violence surrounding the UN-sponsored popular consultation in 1999. We arrive in Mauxiga just before midday and after a quick cup of Timor coffee and some moments taking in the natural beauty of this tiny village, we get to work. Mauxiga is nestled under the watchful eye of Mount Kablaki, amidst some of the most fertile coffee producing and vegetable growing areas of Timor. However, in August 1982 it was the focus of an armed attack by the pro-independence movement against a number of Indonesian army posts. This opposition was met with a strong military response from the Indonesian Army which involved mass arrests, forced displacements and arbitrary detentions of those suspected of contributing to or supporting the resistance movement. It is these human rights violations which Abilio dos Santos, the village secretary of Mauxiga and Olga da Silva, a teacher at the local primary

school have spent the last 19 years meticulously documenting in the hope that the plight of villagers in Mauxiga in the 1982 will not be forgotten or denied by future generations.

As we go through the lists we note some structural ambiguities in the data and how we can represent the data in a way which does not make false simplifying assumptions. In particular, the way the lists had been constructed implies that one victim could only have suffered exactly one violation. So to solve this problem, we restructure the data model to allow for the fact that any given victim may have suffered any number of non-fatal violations in the reference frame². So we work through the lists, and through narrative testimony and some careful thumbing through the original field notes from which the data was constructed Abilio is able to reconstruct the one-to-many relationship of violations, victims and perpetrators which were experienced by the people of Mauxiga.

²For a description of this and other common data representation errors of human rights violation information, refer to P. Ball, *Who Did What to Whom: Planning and Implementing a Large-Scale Human Rights Data Project* (AAAS Washington DC 1996).

From the reworked data, I develop some descriptive statistical analysis of the reported arrests, detentions, displacement and conflict-related mortality in Mauxiga. As I guided Olga and Abilio through the preliminary graphs and analysis which give statistical representation of their data we discussed hypotheses which the data support and possible explanations for the resulting statistical patterns and trends. A significant part of my job is often spent presenting and explaining descriptive statistical analysis or estimates of the extent and pattern of human rights violations to non-technical colleagues. By being able to defer to empirical quantitative data which has been carefully cleaned, edited and developed into a coherent statistical analysis, our partners and colleagues in the human rights movement are able to make scientifically defensible human rights findings against established international human rights norms and concepts. Such notions as whether the extent and pattern of violations are “widespread” (as opposed to following a purely random distribution), “systematic” or follow patterns which are consistent with the hypothesis that such acts were “part of a policy or plan by those responsible” are all measurable and testable when empirical data is available. As Abilio and Olga examine the age-sex pyramids and time-series graphs which I have developed from their data they launch in and out of historical accounts which contribute a rich and detailed context to the meaning of this statistical data. With the expert eyewitness accounts from Abilio and Olga and with their comprehensive empirical data at my fingertips, we are able to review the population-based detention and killing rates as we discuss whether the data supports the argument that such violations were targeted against certain civilian sub-populations or whether such acts appear to have been more indiscriminate.

As we finish working through the queries and clarifications I have about the lists, Abilio and Olga ask about the overall scope

and nature of our statistical analysis on past human rights violations in Timor-Leste. I explain that in partnership with the CAVR, we have created two important innovations that will greatly assist future truth commissions and other large-scale human rights data projects. First, the CAVR created new data sources never before foreseen in work by truth commissions, and second, the CAVR statistical work was on such a massive scale that it required adoption of statistical algorithms for matching repeated reports of the same deaths.

Most truth commissions base their empirical findings principally on databases derived from the large-scale collection of qualitative testimonies. In this, the CAVR was no different from the commissions in Haiti, South Africa, or Peru. In other countries, the truth commissions were able to draw on substantial additional information that had been collected by governmental and non-governmental human rights projects. Additional sources are important in order to “triangulate”, or understand the patterns and magnitude of human rights events from perspectives other than the commission’s own qualitative material. Without outside corroboration, commission’s work could be dismissed as partisan.

The CAVR did not have massive external sources available, so new sources were created. First, the CAVR conducted a retrospective mortality survey of 1396 households that were randomly selected from Timor-Leste’s approximately 180,000 households. Each sampled household gave information about their residence pattern and household members and relatives who died during the CAVR mandate period. Mortality surveys of this kind are common among governmental statistical offices to assess health conditions or to adjust censuses. Inter-governmental health authorities and academic demographers and epidemiologists also conduct surveys of this kind. However,

no truth commission has ever done a rigorously sampled household survey, of any kind.

A third dataset collected by the CAVR was the graveyard census database. Every cemetery in Timor-Leste was visited, and the name, date of birth, and date of death was recorded for every grave for which the information was available. Approximately 327,000 grave records were collected; after duplicate enumerations are removed, there are approximately 319,000 unique graves in the sample, of which about half have complete name and date information. Cemetery records have been used by historical demographers to reconstruct historical patterns of mortality, but no truth commission has ever used data of this kind as part of the reconstruction of historical memory. In the world of human rights measurement, these are tremendous innovations which greatly enrich our understanding of the past.

A second CAVR innovation is purely technical: the adoption of automatching techniques. A common problem in large-scale data collection (of all kinds, not just human rights) is that the same people, deaths, or other events are frequently reported by multiple witnesses or registries. In the CAVR testimonies, for example, the same human rights violations may have been witnessed by several people, each of whom gave statements to the CAVR. It is essential to count each unique event only once so that we do not overcount duplicate reports. Analysis of the duplicate reports can also provide a basis for statistical inferences via multiple systems estimation (MSE) about how many events were never reported at all³. Finding the duplicated reports in the

same dataset and across multiple dataset is called “matching” or “record linkage”⁴.

The CAVR project was faced with much larger datasets than previous projects; the graveyard census, in particular, is vastly larger than any previous truth commission dataset. As a result, the complexity of the matching problem was considerably greater. As a result, my colleagues and I adopted complex techniques from mathematical statistics that assign a probability of matching between every pair of records in the combined data. The “likely” matches were reviewed by a person, but the person was then required only to review likely matches, reducing by nearly ten-fold the amount of human effort required to “de-duplicate” or match the data.

Using this empirical data, my colleagues at HRDAG and I have been developing scientifically-defensible estimates and analysis of the magnitude and patterns of conflict-related mortality (namely civilian killings, disappearances, combatant deaths and famine-related deaths) and non-fatal violations (such as detentions, tortures and displacement). We are analyzing this statistical evidence in light of international human rights norms to test whether human rights violations in Timor were “widespread”, “systematic” or followed patterns which suggest evidence that they were coordinated or part of a formal policy by those responsible for the violence. This analysis is also being integrated with the Commission’s legal, historical and qualitative research.

At present we are now working together with the staff of the CAVR to draw conclusions, based on the newly established empirical data, that are among the bases for the Commission’s most important findings.

³MSE are explained in S.E. Fienberg, M. S. Johnson, and B.W. Junker, *Classical multilevel and Bayesian approaches to population size estimation using multiple lists*, Journal of the Royal Statistical Society, Series A **162** (1999), 383–405. MSE techniques have previously been employed in human rights projects in Guatemala, Kosovo and Peru. For example, see P. Ball, *Policy or Panic: The Flight of Ethnic Albanians from Kosovo, March–May 1999* (AAAS Washington D.C 2000).

⁴There is extensive statistical literature on record linkage methods. See for example P. Christen, T. Churches and M. Hegland, *A Parallel Open Source Data Linkage System*, Proceedings of the 8th Pacific-Asia Conference on Knowledge Discovery and Data Mining, Sydney, May 2004.

These findings address the total extent, pattern, trend and levels of institutional responsibility for past human rights violations in Timor between 1974 and 1999. Our descriptive statistical analysis of arbitrary detentions, forced displacement and conflict-related deaths in Mauxiga will be an important episode about the nature of the conflict in the Eastern region of Timor-Leste in the early 1980's.

As we close our discussions and sit down for some dinner, Abilio quizzes me about how I got involved in quantitative analysis of large scale human rights violations. I explain that in the late 1990's as I was completing my honors degree in statistics at UNSW, I became increasingly involved with Amnesty International both on the campus and at the national level. At the time, Amnesty International Australia was heavily focused on the human rights situation in Timor-Leste. As I spent my time organizing public seminars, campus rallies and petitions calling for greater human rights protections in Timor, I began to wonder how I might be able to directly apply my mathematical education and quantitative skills to international human rights. Having read Nobel Laureate Amartya Sen's writings on extreme deprivation and socioeconomic rights, I started to think about ways in which his elegant measurement framework for human rights might be implemented in war-torn societies or countries which are under authoritarian rule⁵. Upon obtaining my first class honours degree, I went off to Laos to work for the United Nations Development Program (UNDP) developing statistical analysis for their flagship publication on

poverty and rural development. The experience heightened my awareness and understanding of the challenges of collecting data in developing countries. Over the last four years I have focused exclusively on data collection, data processing and statistical analysis of large-scale human rights violations for projects HRDAG projects in Sri Lanka, Timor-Leste and Chad.

My education in mathematics and statistics has helped me adapt quantitative methods from the literature on inter-rater reliability measurement, survey methods and record linkage to develop scientifically defensible measurements of large scale human rights violations. As the field of human rights statistics is relatively new, our main challenge remains in attracting more mathematicians and statisticians to work in this area and also communicating to the human rights field how quantitative data analysis can contribute to accountability and truth-telling processes.

The Commission will hand over its findings to the President of Timor-Leste in June 2005. At this time, the Commission's Final Report which will span more than 2,000 pages and be published in English, Portuguese, Tetun and Bahasa-Indonesia will be released publicly. HRDAG will simultaneously publish a series of additional, in-depth methodological publications on the Benetech-HRDAG web site. In addition, HRDAG will publish the Commission's statistical data (stripped of all personally-identifiable information) in order to promote ongoing research and further discussion between the scientific and human rights communities on human rights accountability for past atrocities in Timor-Leste.

E-mail: Romesch.S@benetech.org

⁵In particular, for me the two most influential ideas of Sen's were (i) his conceptualization of human development and basic rights within a framework of human capabilities, and (ii) his lucid framing of the poverty measurement as a problem of "identification" and "aggregation". For a more detailed discussion of these ideas see A. Sen, *Sen, Poverty and Famines; An Essay on Entitlement and Deprivation* (Oxford Clarendon Press 1981).