
MATCHING THE *Libro Amarillo*
TO HISTORICAL HUMAN RIGHTS DATASETS
IN EL SALVADOR

Patrick Ball
August 20, 2014



1 Introduction

This project considered 1857 records identified by name in the *Libro Amarillo*, the Yellow Book. We compared the Yellow Book names to four databases that were organized in El Salvador: the non-governmental Human Rights Commission of El Salvador (CDHES); the published reports of the Tutela Legal del Arzobispado organized by El Rescate (Rescate); the database of violations reported to the truth commission through direct testimonies (CV1); and the database of violations reported by the truth commission as “indirect sources.”

The intersections of the Yellow Book with the other datasets is shown and described in Section 1.1 and shown in Table 1.

This report makes no substantive conclusions about the patterns. It only describes the quantitative patterns of links among the Yellow Book and the other lists.

Numbers in this report are formatted according to the [International Standards Organization 31-0](#). The thousands separator is a small space, and the decimal is marked by a point. Technical jargon and Spanish words are shown in *italics*; the names of specific computer programs, variables and fields used in the matching are shown in `typewriter`. Links within the document are shown in red, and links to external websites are shown in blue.

1.1 Findings

Table 1 shows the numbers of names found in the Yellow Book that can be matched to to names in the other lists.

Table 1: Record Counts and Overlaps Between Historical Datasets and the *Libro Amarillo*

	killling	disappearance	torture	detention	other
CDHES	35	75	157	337	179
Rescate	93	65	43	397	66
CV1	127	21	7	NA	7
CV2	124	151	123	NA	6
YB	273	233	274	538	242
YB percent	15	13	15	29	13

This means that, for example, 127 names from the Yellow Book were matched to a name recorded as a killing in the truth commission’s direct sources (CV1), while 21 names from the Yellow Book were matched against names who were disappeared in the truth commission’s direct sources. (Note that neither of the truth commission sources reported detentions.)

The same victim may have suffered both fatal and non-fatal violations. The numbers in Table 1 show the numbers of *violations*, not victims. The same victim may have suffered one or several violations, and therefore, the rows cannot be summed across.

Similarly, the same name from the Yellow Book might be found in one or several databases. Therefore the columns of Table 1 cannot be summed. Instead, the count of names in the Yellow Book summed over all the datasets is shown under the penultimate line. For example, the count of names that match to killings documented in one or more of the other lists is 273, and the total number of names that match to a disappearance record in the Yellow Book is 233. The percent of names in from the Yellow Book matched to one or more records in the other lists is shown in the final rows: 29% of the names in the Yellow Book are matched to at least one record of a detention.

Lastly, 1063 records from the Yellow Book were not found in any of the other lists.

Table 2: Counts of Violations Against People Named in the *Libro Amarillo* Attributed to Armed Institutions in the Historical Datasets

	killings	disappearance	torture	detention	other
Army	177	71	109	208	81
Security	70	73	160	344	68
Death Squad	128	43	32	81	37
FMLN	19	12	0	1	2
YB	273	233	274	538	242
YB percent	15	13	15	29	13

The names from the Yellow Book linked to violations documented in the other lists (i.e., the last two rows of Table 1) can be subdivided by the perpetrator to which the violation was attributed in the original documentation. Table 2 describes violations suffered by people named in the Yellow Book with the perpetrators alleged to have committed the violations. For example, the first cell of Table 2 shows that according to the other lists, 177 people named in the Yellow Book were

killed by the Salvadoran Army, while 128 people named in the Yellow Book were killed by death squad groups. Note that the rows of Table 2 cannot be summed. The same victim may have been killed by an operation in which both the Army and death squad units operated, so that killing would be counted in both of those rows in Table 2. Similarly, the same victim might have suffered both killing and torture, and so the columns cannot be summed. The final two rows of Table 2 are repeated from Table 1.

To calculate the percent of people named in the Yellow Book who suffered a given violation by a given perpetrator, divide the count from the relevant cell of Table 2 by the total number of people named in the Yellow Book and considered in this study (1857). Thus 10% of the people named in the Yellow Book were killed by the Army, while 19% of the people named in the Yellow Book were detained by the security forces.

2 Methods

We *matched* each name in the Yellow Book to the eight *closest* names in the combined set of names from the *historical databases*. We define the emphasized ideas in reverse order:

The *historical databases* include the CDHES;¹ the database of the El Rescate/Tutela Legal project;² the truth commission’s direct list (CV1); and the truth commission’s indirect list (CV2).³ Together these are referred to as the *other lists*, and they contain 62 292 records.⁴ Note that a Yellow Book name might match to zero, one, two, three, or all four of the other lists.

By *closest*, we mean names that are most similar to each other. One way to imagine name similarity is to ask how many characters from one name would need to be inserted or deleted in order to convert it into the other name. That measure is called [Levenshtein distance](#). In this project, we used a related measure called the [Jaro score](#) ([Winkler, 2006](#)).

The Jaro score is optimized to test the similarity of names, and it was developed by statisticians at the US Bureau of the Census. It has a standardized

¹See [Ball \(2000\)](#).

²See [Howland \(2008\)](#).

³See [United Nations \(1993\)](#).

⁴Among the 62 292 records, there were 37 520 unique names because some records had exactly the same names as other records.

range between 0 and 1. I tested each of the 1857 names in the Yellow Book against all of the 37 520 unique names⁵ found in one or more of the other lists.

That is, we calculated the Jaro score for each pair of names from the 1857 names from the Yellow Book and the 37 520 unique names from the other lists. For each Yellow Book name, we chose the eight most similar names from the other lists, that is, the eight names with the highest Jaro scores; a person then reviewed those eight comparisons. In a few cases, there were fewer than eight remotely plausible names in the other lists to compare to the Yellow Book name (remotely plausible’ means a Jaro score > 0.7). In total, there were 14 833 combinations of names for human review.

Before we made the comparisons, we sorted the *name tokens*, where token means a piece of a longer name. This means that if one name was written “CACERES JOAQUIN” and the other was written “JOAQUIN CACERES CERES,” we would sort each name string in order to compare “CACERES JOAQUIN” to “CACERES CERES JOAQUIN.” The resulting Jaro score is 0.905, indicating that the two names are quite similar.

By *match*, we mean that a human being reviewed by hand 14 833 combinations of names. Many of the comparisons were made twice by two different people, and the comparison of their findings is described in Section 4.1.

3 Data Appendix

Table 3 shows the total numbers of records in the other lists.

Table 3: Record Counts in the Historical Datasets of Human Rights Violations in El Salvador

	killings	disappearance	torture	detention	total
CDHES	1870	2555	3708	8745	24 550
Rescate	5030	3040	950	6010	16 929
CV1	5604	994	434	NA	7258
CV2	7376	3682	2230	NA	13 555

⁵Many names were in the other lists more than once; the identical names from different records in the other lists were combined for the comparison to the Yellow Book names.

4 Measurement Appendix

Matching is fundamentally about deciding whether a pair of records does or does not refer to the same person. Given a pair of names, people doing the matching may make the same decision, or different decisions. The measures described here explain how we assess the similarity between two different matchers' decisions.⁶

In the context of this study, one matcher was arbitrarily chosen as A and the other as B. With those assignments, we define the following

- a *true positive* (TP) is a pair that is labeled by both people as a match;
- a *false positive* (FP) is a pair that is labeled by person A as a non-match but by person B as a match;
- a *false negative* (FN) is a pair that is labeled by person A as match but by person B non-match;
- a *true negative* (TN) is a pair that is labeled by both people as a non-match.

Given these categories, we can define three additional measures. First is *recall*, the proportion of all matches that are classified properly. The recall value describes the rate at which B is saying non-match while A is saying match. A higher recall score means that there is less disagreement of this kind.

$$recall = \frac{TP}{TP+FN}$$

Second is precision, which describes how often B is saying match while A is saying match. A higher precision score means there is less disagreement of this kind.

$$precision = \frac{TP}{TP+FP}$$

The *F-measure* is a balance between precision and recall, and is defined as follows.

$$F = \frac{2 \times recall \times precision}{recall + precision}$$

F is a generalized measure of the quality of classification. The F -measure is used to describe the inter-rater reliability in Section 4.1.

⁶These measures are used widely in the statistical and computer science literature on machine learning. See, for example, [Witten et al. \(2011\)](#), especially chapter 5.

4.1 Inter-Rater Reliability

The reliability of the matching result is fundamentally based on whether the decisions people made about whether two names are, or are not, the same person are correct. It is impossible to evaluate the true correctness—we do not know exactly who each person is in the Yellow Book, or for that matter, in the other lists. Instead, we consider how often two people looking at the same pair of names decide that they are the same person versus how often they disagree. The two matchers reviewed a total of 1604 pairs in common. The measure of their agreement or disagreement is called *inter-rater reliability* (IRR). There are several ways to measure IRR, but for this project, we will use the F measure defined in Section 4.

For matches, $F^+ = 0.79$, while for non-matches, $F^- = 0.91$. In essence, the human raters agreed about non-matches, but they agreed less often about positive matches.

The ideas of *precision* and *recall* provide insight into the kind of disagreements the matchers had. For positive matches, $precision^+ = 0.66$, $recall^+ = 0.99$. The relatively weaker score for precision suggests that the matchers tended to disagree such that one of them more frequently said that a pair of names was a match. The relatively high recall score suggests that the disagreement was one-sided: one of the matchers was simply more reluctant to identify the pair as a match. However, when the reluctant matcher said the name pair was a match, the more aggressive matcher nearly always agreed.

For the purposes of this study, the more reluctant matcher’s decisions were used, specifically in order to estimate a minimum level of overlap between the Yellow Book and the other lists.

References

- Ball, P. (2000). The Salvadoran Human Rights Commission: Data Processing, Data Representation, and Generating Analytical Reports. In Ball, P., Spierer, H. F., and Spierer, L., editors, *Making the Case: Investigating Large Scale Human Rights Violations Using Information Systems and Data Analysis*. AAAS, Washington, DC.
- 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):703–757.
- United Nations (1993). *From Madness to Hope: The 12-Year War in El Salvador: Report of the Commission on the Truth for El Salvador*.
- Winkler, W. E. (2006). Overview of Record Linkage and Current Research Directions. Technical Report RRS2006/02, Statistical Research Division, U.S. Census Bureau.
- Witten, I., Frank, E., and Hall, M. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufman.

About HRDAG

The [Human Rights Data Analysis Group](#) is a non-profit, non-partisan organization⁷ that applies scientific methods to the analysis of human rights violations around the world. This work began in 1991 when Patrick Ball began developing databases for human rights groups in El Salvador. HRDAG grew at the American Association for the Advancement of Science from 1994–2003, and at the Benetech Initiative from 2003–2013. In February 2013, HRDAG became an independent organization based in San Francisco, California; contact details and more information are available on HRDAG’s website (<https://hrdag.org>) and [Facebook page](#).

HRDAG is composed of applied and mathematical statisticians, computer scientists, demographers, and social scientists. HRDAG supports the protections established in the Universal Declaration of Human Rights, the International

⁷Formally, HRDAG is a fiscally sponsored project of [Community Partners](#).

Covenant on Civil and Political Rights, and other international human rights treaties and instruments. HRDAG scientists provide unbiased, scientific results to human rights advocates to clarify human rights violence. HRDAG believes that statistics about violence need to be as true as possible, using the best possible data and scientific methods.

This project was designed, the software written, and all processes overseen by [Patrick Ball](#). Technical ideas were developed in collaboration with [Dr. Scott Weikart](#) and [Dr. Megan Price](#). The hand matching was done by HRDAG data management consultant [Michelle Dukich](#), and by Phil Neff, Unfinished Sentences Project Coordinator, University of Washington Center for Human Rights.”

The materials contained herein represent the opinions of the authors and editors and should not be construed to be the view of HRDAG, any of HRDAG’s partners, the HRDAG Board of Advisers, the donors to HRDAG or to this project. “Matching the Libro Amarillo to Historical Human Rights Datasets in El Salvador” by Patrick Ball is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](#). Permissions beyond the scope of this license may be available at <https://hrdag.org>.

