Different Convenience Samples, Different Stories:

The case of Sierra Leone

by Anita Gohdes*

April 6, 2010

Benetech

1 Summary

This analysis examines the differences and similarities of the three data sources collected during and in the aftermath of the conflict in Sierra Leone. In particular, it attempts to examine the relative frequency of reported violations across the three datasets for relevant strata such as the year the violation occurred, sex and age of the victim, as well as ethnic and regional differences. The comparison of these rankings offers a non-parametric, scale-invariant and dimensionless method for the analysis of the different reporting patterns between the data sources at hand. The results presented here reveal that even for a relatively simple measure of similarity such as rank correlations, the structure of reporting differs between all three sources. The findings support the hypotheses that single sources of information of conflict situations only capture a part of the overall pattern of violence, that this information is almost always biased, and that there is no way to assess this bias using a single dataset.

2 Objective

Past experiences of analyzing patterns of violence in conflict situations have taught us that information gathered on the occurrences of human rights violations by any single source can be biased in different ways. The reasons for this are manifold and range from problems of accessibility and willingness of the individual to report the witnessed or endured violations, to financial, political and social variations of the institution that is gathering the data, and lastly characteristics of the violation itself that make it more or less susceptible to reporting (see Hoover et al., 2009). Equally, we have learnt that triangulating different data sources by using statistical modelling and estimation can help us get a more comprehensive picture of conflict situations (see, e.g. Ball et al., 2003, Ball and Silva, 2007; Guberek et al., 2010). The quality of this picture is, however, dependent on whether we are able to take into account the potential biases of the individual sources and include these in our estimation process. In order to do this, we need to understand the problems and failures of the individual sources to be included in the statistical model. The question arises of how we can systematically analyze the differences or similarities in reporting patterns across different sources of information. Do these datasets tell us different stories about the same conflict? And if so, where do we begin depicting these stories?

3 Description of the three data sources

The three data sources of interest are the ABA/Benetech Sierra Leone War Crimes Documentation Survey (SLWCD), the dataset created out of the statements of violations taken by the Truth Commission (TRC)[1], and those collected by the non-governmental organisation 'Campaign for Good Governance' (CGG). For the SLWCD dataset, the present analysis uses the unweighted raw violation counts of the survey, and reports the weighted totals in a separate ranking, since they were estimated for the overall population of Sierra Leone. The data collected by the Truth Commission and the Campaign for Good Governance both follow human rights database design standards. The information provided by the Sierra Leone War Crimes Documentation (SLWCD) database follows

---

[1]The data collected by the TRC is available at: http://hrdag.org/resources/SL-TRC_data.html.

general household survey questionnaire design standards.[2] All three data sources present exemplary cases of data collection efforts. The fact that three sources of information of such high quality are available for the entire conflict period is rare and gives us the opportunity of exploring the characteristics of different samples of reports over the same time and space.

4 Correlations of ranked violation types

This paper looks at the relative differences in the frequency of reported violations. By examining the six most frequently reported violations across all three datasets, the changes in the relative frequency of these violations between the different sources gives us an idea of how similar or dissimilar the reporting patterns are.

The violation types are ranked for each source individually, according to the frequency with which they were reported. In order to obtain a measure of (dis)similarity of sources we look at the differences in overall rankings, as well as when stratified by different time periods, by sex, by age groups, for different regions and by ethnicity of the victims.

For each of these strata, we calculate Spearman's rank correlation between the different pairs of data sources (SLWCD and TRC, SLWCD and CGG, TRC and CGG), which in turn leaves us with three correlation coefficients that provide us with a measure of association of the pairs of data sources. Since the household survey data does not include estimations (i.e. weighted observations) for the violation type "killing," the overall correlation of the weighted survey data with the other two sources is calculated separately. The three correlation coefficients are visualized in a triangle with the help of radar graphs.

---

[2] A 'case' is defined as the information given by a single deponent concerning violations that happened at a particular time and place.
'Violations' are instances of violence, including killings disappearances, torture, acts of displacement and acts of property destruction.
'Victims' are people who suffer violations. A human rights 'case' may be very simple (with one victim who suffered one violation) or it may be very complex (with many victims each of whom suffered many different violations) (see Guberek et al., 2006: 5).

Spearman's rank correlation coefficient ($corr$) offers a simple and yet effective measure for the comparison of the structure of the data sources. The coefficient offers information on whether the pair of datasets under scrutiny has exactly the same ranking ($corr = 1$), has an inverse ranking of violations ($corr = -1$) or differs somewhere in between these extremes ($-1 < corr < 1$). By using a ranking measure, we are able to assess structural differences without having to take into account differences in size or dimension of the data. For example, let's assume that we, hypothetically, knew that throughout the Sierra Leone conflict many more people suffered property theft than were assaulted, and that the third largest problem the population faced was displacement from their homes. In such a situation we would assume that all of the data sources should reflect this "order" of violations, and that when ranked, Spearman's correlation coefficient between the pairs of sources should approach $corr = 1$. The triangle that represents these coefficients within the radar graph should thus span out widely, with the three corners touching the outer circle of the graph, denoted as 1.

Evidently, we never find ourselves in the situation of knowing exactly what happened, and the above-mentioned challenges of recording violence can lead to many different forms of bias. Some people might be more inclined to report property violations to the Truth Commission, in the hope of receiving retribution. In such a case, destruction of property violations would rank higher in the TRC's dataset. Equally, killings might be reported irregularly to different sources and across the different strata. Depending on where and when a person was killed, it is possible that the person reporting this violation might only be willing to testify, if at all, to one organisation, since multiple reports could be re-traumatizing or logistically speaking, too much effort.

These examples demonstrate that the reported frequencies of violations may vary due to characteristics of the organisation gathering the information, due to victim characteristics, due to exogenous differences, such as time and location and lastly, due to differences associated with the types of violations themselves. Many of these factors lead to biases that cannot be measured or observed in a straightforward way. The following analysis attempts to assess differences for some of the characteristics that *can* be measured.

4.1 Overall

Table 1 lists the six most reported violations across all sources and their respective ranking for each of the three datasets. Forced Displacement is ranked as the most frequently reported violation across all sources. The Household survey (SLWCD) and the data collected by the Campaign for Good Governance (CGG) are relatively similar in that they share the first three ranks. The data collected by the Truth Commission (TRC) shares the first and third rank with the other two sources, but overall, follows a slightly different pattern. Whereas the SLWCD and CGG rank Assault/Beating as the second most frequently reported violation, the Truth Commission only ranks this category of violation fifth, and Arbitrary Detention second.
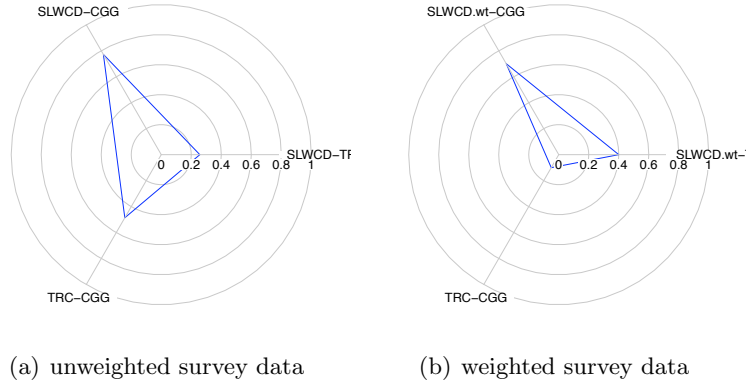
Table 1: Overall Ranking of Top Six Violations

Violation	SLWCD	CGG	TRC
Forced Displacement	1	1	1
Assault/Beating	2	2	5
Destruction of Property	3	3	3
Killing	4	6	6
Arbitrary Detention	5	5	2
Property Theft	6	4	4

Table 2: Overall Ranking of Violations, excluding killings

Violation	SLWCD	TRC	CGG	SLWCD.wt
Forced Displacement	1	1	1	1
Destruction of Property	3	3	3	2
Property Theft	5	4	4	3
Assault/Beating	2	5	2	4
Arbitrary Detention	4	2	5	5

Table 2 includes the overall rankings of the weighted survey data. Since no weights could be calculated for the violation type "killing," the ranking is conducted for the remaining five most frequently reported violations only. As we can see there is a considerable difference in ranking between the weighted and unweighted survey data, which is reflected in the very different rank correlations between the datasets.

Figure 1: Overall Rank Correlations

(a) unweighted survey data

(b) weighted survey data

The radar graph in Figure 2(a) visualizes the three rank correlation coefficients for each data-pair as presented in Table 1. The size and shape of the triangle within the graph represents the way in which the rankings of the datasets match or do not match. For example, in Figure 2(a), the corner of the triangle that represents the correlation of SLWCD and CGG spreads all the way to the circle that is labelled with 0.8, which means that the Spearman rank correlation of these two datasets is almost $corr = 0.8$. In comparison, the overall rank correlation between the TRC and CGG is much smaller ($corr < 0.5$). Lastly, the corner representing the relationship between the household survey data and the TRC is the smallest, with a correlation of less than $corr = 0.3$.

Figure 2(b) visualizes the correlations that were calculated from the ranks in Table 2. When compared to Figure 2(a), we see that the patterns of violation types are still relatively similar between the NGO (i.e. the CGG) and the survey data (SLWCD), but that without the relative frequency of reported killings, the NGO and TRC data now seem to coincide even less with each other.[3]

---

[3]Note that for this reason the unweighted and weighted correlations can strictly speaking not be compared since a large proportion of the difference in structure results from the omissions of this important violation type.
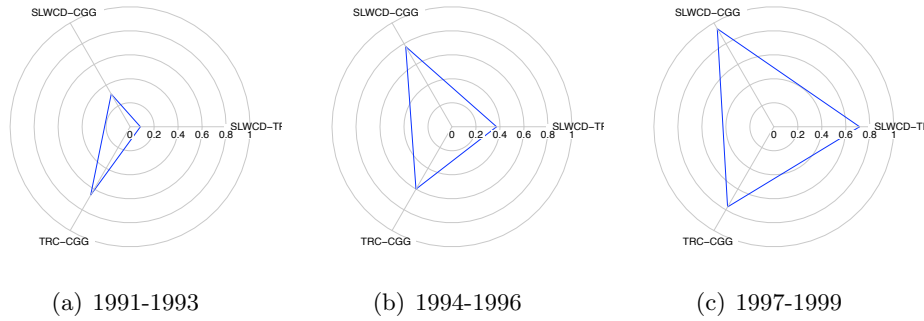
4.2 By Time Periods

According to the report published by the Truth Commission, the Sierra Leone civil war can roughly be divided into three main conflict phases. This claim is reflected in all three data sources, where peaks of violence are visible in 1991, 1994/1995, and 1998/1999. For the purpose of this analysis, a ranking of the six most frequently reported violations is therefore undertaken for the three time periods ranging from 1991 to 1993, from 1994 to 1996 and from 1997 to 1999.[4] Figure 2 displays how the similarity in rankings significantly increases over time. For the time period between 1991 and 1993, the correlation of the relative reported violation frequencies is relatively low, especially when comparing the household survey data with the other two sources. The correlation coefficient for the period from 1993 to 1996 almost triples for the survey and the NGO data, and the other two coefficients equally increase. For the last peak conflict period, the three data sources support a similar picture with regard to the relative frequency of reported violations.

What does this discrepancy between reporting patterns of the three data sources at the beginning of the conflict tell us about the reliability of the data sources? There are multiple hypotheses that would support this evidence. For example, this finding might indicate that people's memories of the conflict are more variable in the earlier than in the immediate past. This could imply that people tend to agree on the immediate past but recall events in the further past differently. These changing correlation coefficients might be an indicator for a potential recall bias that would affect the validity of evidence gathered retrospectively. Finally, the differences across time might point to the fact that the three data sources were capturing different sectors of the Sierra Leone society at the beginning of the conflict, but had access to similar sectors for the later years. In any case, these findings display how each individual dataset offers a different version of what happened in the early years of the civil war.

---

[4]For the remaining years of the conflict the data is too thinly spread to make valid conclusions on the congruence of the relative reported frequencies.
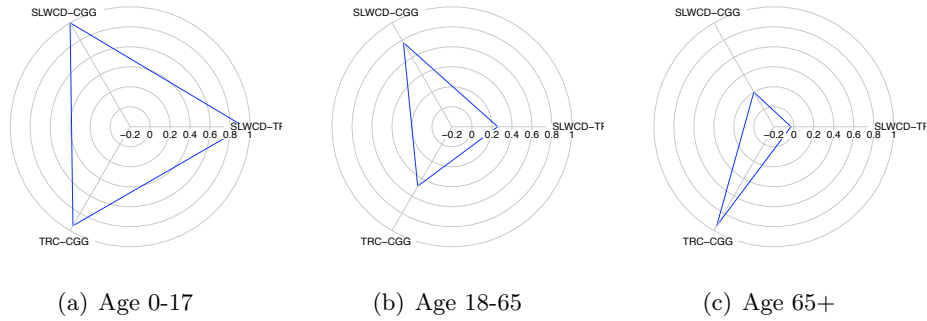
Figure 2: Correlations, by Time Periods

(a) 1991-1993   (b) 1994-1996   (c) 1997-1999

4.3 By Age Groups

The ranking for all three sources is more similar in the case of younger victims. The rank correlation coefficients decrease with increasing victim age. Again, different explanations are possible: Are all three sources accessing a similar sector of society when it comes to the reports of young victims? Are young victims generally more visible, or are only specific violations in this age group more susceptible to reporting? In the same line of arguments, one could assume that older victims might be more reluctant to come forward and speak about past abuses. In the case of personal integrity violations, older members of the population that suffered abuses earlier on in the conflict might have passed away as a consequence of these injuries. Turning to the characteristics of the different data sources, older members of society might be more reluctant to repeat their story, or the story of their peers to multiple institutions, and, depending on their political and logistical position, might choose to report selectively. Setting aside the possible explanations for this trend, the differences in congruence of the three datasets with increasing recorded age of the victims demonstrate that a single source of information is not representative of the actual dimensions of violence that occurred.
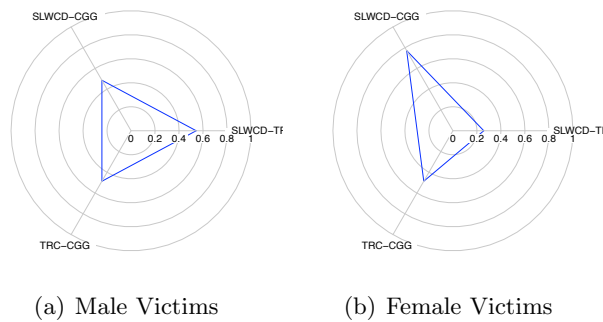
Figure 3: Correlations, by Age Groups

(a) Age 0-17   (b) Age 18-65   (c) Age 65+

4.4 By Sex

It is interesting to note the differences between the overall correlation coefficients and those that were stratified by sex. The female stratum and the overall data seem to have a similar pattern, in that the SLWCD and CGG data coincide more in their ranking of reported violations than the other two data-pairs do. This is particularly interesting in view of the fact that only approximately 35% of all reported violations occurred to female victims. The stratification across sexes shows us that the data collected by the three databases follows a different structure for male victims than it does for violations that were reportedly suffered by women.
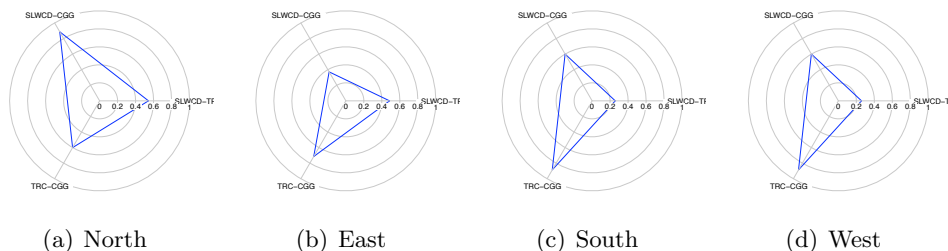
Figure 4: Correlations, by Sex

(a) Male Victims   (b) Female Victims

4.5 By Location of violation

How does the similarity in reported violations vary across geographic regions? Figure 5 gives us an overview of the correlation coefficients for the six most frequently reported violations. For the South, West and East of the country, the data collected by the TRC and the CGG follow a more similar pattern than when compared with the survey data. The violations that were recorded to have occurred in the North of the country are overall more congruent than the other regions, which might be explained by the fact that 40% of all violations were reported here, and only 31% were reported in the East, 23% in the South and only 6% in the West. It could also be an indication that all three sources collected the information on the same abuses in the North, and that the TRC and the CGG recorded more of the same statements in the rest of the country than the household survey did.

Figure 5: Correlations, by Location of Violation

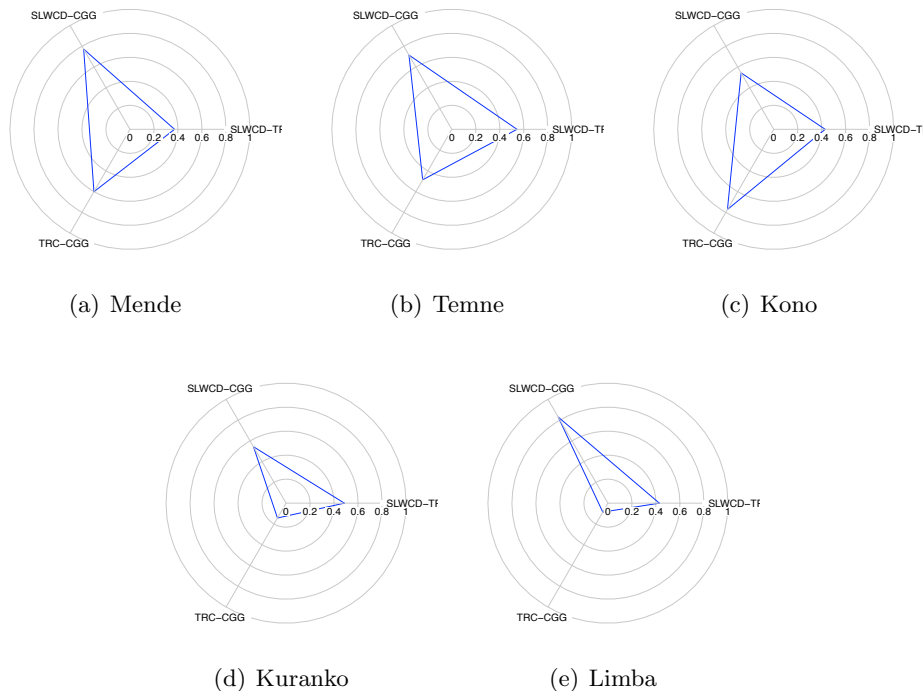(a) North   (b) East   (c) South   (d) West

4.6 By Ethnicity

Finally, we look at the five most frequently reported ethnicities of victims. The rank correlations for the three data-pairs vary significantly for the different ethnicities. Not only does the ranking of violations change between the different ethnicities, the order of the rank correlations also changes. For example for those victims identified as Kono, the data collected by the TRC and the CGG ranks in a very similar way, but for the

other four ethnicities analyzed here, the structure of these two datasets seems to be very different. It is interesting to note that the five most frequently reported ethnicities of victims do not constitute the five largest ethnicities within Sierra Leone. The first two groups seem proportionally represented, as the Mende and Temne each make up roughly one third of the country's population. In contrast, the Kuranko rank as forth most frequently reported ethnicity, even though they only rank eighth overall. The Limba, who are the third largest group within the country, only rank fifth for the most reported ethnicities. For the both the Kuranko and the Limba, the data collected by the Truth Commission and the CGG seem to tell completely different stories, as their correlation coefficient is almost zero. Does this mean that some groups within society, in this case ethnicities, are captured more uniformly and completely than others? And is the variation in correlations of the same datasets across different ethnicities an indicator of different levels of trust between the data-gathering institutions and the members of different groups?

Figure 6: Correlations, by ethnicity of the victim

(a) Mende   (b) Temne   (c) Kono

(d) Kuranko   (e) Limba

5 Conclusion

This analysis has attempted to find an entry point for systematically looking at differences between the three data sources that were collected for the same period of time on the violations committed in Sierra Leone. In comparing the relative reported frequencies of violation types, the findings show that the information gathered by the three sources differs not only between the datasets as a whole, but that these differences are not consistent and only show limited structure when stratified across important victim traits and time periods. It is particularly notable that there seems to exist a significant change in congruence of the data sources when looking at different time periods. The findings presented here indicate that statements regarding violations that occurred at the beginning of the conflict follow a very different pattern to the violations that were reported to have occurred towards the end of the war, a conclusion that casts considerable doubt on the reliability of data collected about events that occurred in the further past. Equally, it is interesting to see that the information gathered on younger victims shows much stronger similarities than for victims who were adults at the time of the incidence, and even less so for those aged 65 or above.

These preliminary findings offer a first indication of how individual databases can be biased and can thus not be understood as a representative reflection of past events: The accessibility of statement givers and the information they offer is non-random and biased, which means that none of these sources are, in themselves, representative. The changing levels of congruence when stratified across important characteristics confirm previous analyses that have cautioned the reliance on individual sources as a means to retrospectively inquire about violence (see Hoover et al., 2009). Since the level of agreement varies significantly for *all* strata, we can assume that all of the factors that were listed as potential reasons for bias in the introduction have an effect on the structure of these databases. There most likely exist many more influences that increase the level of discrepancy in the information collected. Whereas the matching of the observations

between the different datasets might help understand to what extent different realities exist, the number of matched cases is unlikely to be high enough to fundamentally change the outcome of this analysis

More in-depth study of reporting patterns, accompanied by qualitative information on the reporting procedures of the different sources is needed to better understand the shortcomings and biases of datasets if they are to be used for the analysis of patterns of violence in situations of conflict.

6 References

Ball, Patrick, Jana Asher, David Sulmont and Daniel Manrique (2003) 'How many Peruvians have died? An estimate of the total number of victims killed or disappeared in the armed internal conflict between 1980 and 2000'. Report to the Peruvian Truth Commission for Truth and Justice (CVR).

Guberek, Tamy, Daniel Guzmán, Romesh Silva, Kristen Cibelli, Jana Asher, Scott Weikart, Patrick Ball, and Wendy M. Grossman, 'Truth and Myth: Human Rights Violations in Sierra Leone, 1991-2000.' A Report by the Benetech Human Rights Data Analysis Group and American Bar Association's Central European and Eurasian Law Initiative. 28 March 2006.

Hoover, Amelia, Romesh Silva, Tamy Guberek, and Daniel Guzmán (2009)' The "Dirty War Index" and the Real World of Armed Conflict'. 24 March 2009.

Guberek, Tamy, Daniel Guzmán, Megan Price, Kristian Lum and Patrick Ball, 'To Count the Uncounted: An Estimation of Lethal Violence in Casanare.' A Report by the Benetech Human Rights Program. 10 February 2010.

Silva, Romesh and Patrick Ball (2007) 'The Demography of Conflict-Related Mortality in Timor-Leste (1974-1999): Empirical Quantitative Measurement of Civilian Killings, Disappearances & Famine-Related Deaths', in *Statistical Methods for Human Rights*, J. Asher, D. Banks and F. Scheuren (eds), Springer: New York.

The materials contained herein represent the opinions of the author and should not be construed to be the view of the Benetech Initiative. The interpretations and conclusions are those of the author and do not purport to represent the views of the Benetech Board of Directors, any of Benetechs constituent projects, or the donors to Benetech.

This paper should be cited as:
Anita Gohdes. 2010. 'Different Convenience Samples, Different Stories: The Case of Sierra Leone.' The Human Rights Data Analysis Group at Benetech.

Contact information:
The Benetech Initiative
http://www.benetech.org
tel: +1 650-644-3400
fax: +1 650-475-1066
email: info@benetech.org