

Truth and Myth in Sierra Leone:
An Empirical Analysis of the Conflict, 1991-2000

Tamy Guberek, Daniel Guzmán, Romesh Silva, Kristen Cibelli,
Jana Asher, Scott Weikart, Patrick Ball and Wendy M. Grossman

A Report by the Benetech Human Rights Data Analysis Group and the American Bar Association



28 March 2006

1 Introduction

The work presented here adds to that of The Sierra Leone Truth and Reconciliation Commission (TRC) by analyzing data not available at the time of the TRC Final Report publication. Specifically, this document relies on data from three sources: the original TRC data, data collected independently by the Campaign for Good Governance (CGG), and data obtained via the ABA/Benetech Sierra Leone War Crimes Documentation Survey (SLWCD). The TRC did not and could not make arguments about what happened in Sierra Leone beyond the limits of what was reported to it by voluntary deponents. It had no way of measuring the reporting bias inherent in its data collection process; as a result, estimates of overall counts of violations were not possible. The current work uses the two additional datasets that were not available to the Commission, and attempts to assess this bias and estimate the total fatal violations using multiple systems estimation. Additionally, using the SLWCD Survey, the current work makes estimates of the total magnitude of non-fatal human rights violations during the Sierra Leone conflict.

1.1 Objectives of Analysis

This report is preliminary: although the data are now nearly complete, the analysis has many avenues to explore before the estimates findings can be considered definitive. The report uses quantitative analysis and historical analysis, to tell the story of conflict-related violence in Sierra Leone between 1991 and 2000. Statistical analysis helps to identify the pattern and magnitude of large-scale human rights violations in Sierra Leone and to establish the demographic profile of the victims of these violations.

1.2 Summary of Principal Findings

This report has three goals: first, to clarify a number of findings made by the Sierra Leone Truth and Reconciliation Commission; second, to extend the analysis begun by the Commission in order to engage a number of questions and hypothesis which could not be examined by the Commission; and third, to establish the proportional responsibility for the overwhelming majority of conflict-related violence in Sierra Leone between 1991 and 2000.

This report verifies that the Revolutionary United Front (RUF) was responsible for the majority of killings, property violations, and acts of forced displacement. Although the RUF was not responsible for all violations all the time, the RUF was responsible for more violations than any other single group in all years between 1991 and 2000. According to estimates based on the SLWCD Survey, the RUF were responsible for approximately 40 percent and ‘Rebels’ were responsible for approximately 35 percent of human rights violations during the conflict. There is much speculation whether and to what extent the Rebels and RUF collaborated. Other institutions bear much less responsibility for human rights violations: the Armed Forces Revolutionary Council (AFRC) for 11 percent of violations, the Civil Defense Force (CDF) for 4 percent, and the Sierra Leone Army (SLA) for 3 percent.

The nature of violence during the Sierra Leonean conflict was episodic. When violence was large-scale, it tended to be concentrated in particular regions such as Freetown during the 1999 invasion. By contrast, when violence was being perpetrated on a much smaller scale, it tended to be dispersed over a wide geographic space. Over the course of the conflict, the violence spread from being concentrated in a few regions of Sierra Leone to being perpetrated throughout all regions of the country.

The most common forms of violence experienced during the conflict were acts of forced displacement, property destruction, and theft. Of a total population of approximately 5 million, an estimated 1.5 million (+/-145,000) people suffered displacement and 430,000 (+/-35,000) people suffered property destruction. The households that reported suffering displacement suffered approximately 50 percent more property destruction relative to the households that did not report suffering displacement, suggesting that property destruction was a risk factor for displacement.

The demographic profile of victims varied noticeably in relation to the method of violence. The raw data¹ suggest that sexual slavery and rape were targeted specifically against females between the ages of 15 and 19. In contrast, the raw data suggest that males between the ages of 15 and 39 experienced roughly the same relative risks of suffering acts of forced recruitment.

1.3 The Data

The TRC and CGG datasets follow human rights database design standards. The Sierra Leone War Crimes Documentation (SLWCD) database follows general household survey questionnaire design standards. 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). In almost all of the statistics in this report, the unit being counted is the violation.

The following analysis combines information management techniques with statistical analysis and estimation techniques to analyze large-scale empirical human rights information on Sierra Leone in order to inform relevant human rights policy questions about pattern, magnitude and levels of responsibility.

1.4 Magnitude of the conflict

An essential component of statistical estimation is to determine the total magnitude of a phenomenon. This study has benefited from having two estimation mechanisms available. The SLWCD survey enables estimates of the range of non-fatal violations, while the combination of datasets enables estimation of killings.

The estimation technique used to combine several independent samples of information in order to estimate totals is called multiple systems estimation (MSE). Work on MSE estimates for mortality due to violence in Sierra Leone during the conflict is at a preliminary stage. These estimates are not estimates of deaths ‘due to the conflict,’ or excess mortality resulting from all causes. Instead, these are specifically deaths of non-combatants that were intentionally caused by a perpetrator. Most analyses of the Sierra Leone conflict estimate the total killings during the conflict at anywhere from 50,000² to 75,000.³ However, preliminary analysis of the three datasets using multiple systems estimation

¹The analysis presented here uses the raw (unweighted) counts of violations reported to the SLWCD. The patterns in the counts were checked against the weighted estimates, and no substantial differences were found.

²Elizabeth M. Evenson, ‘Truth and Justice in Sierra Leone: Coordination Between Commission and Court,’ p4.

³Ian Smillie, ‘Dirty Diamonds and Civil Society’.

indicates that the range of estimates is between 10,000 and 30,000 noncombatant deaths as the result of intentional violence between 1991-2000.⁴ With additional research, we expect to narrow this range.

The SLWCD survey enables an analysis of non-fatal violations. The most common violations were forced displacement, property destruction, and property theft. An estimated 1.5 million (+/- 145,000) people suffered displacement and 430,000 (+/- 35,000) people suffered property destruction. Other violations were relatively less common. There were an estimated 146,000 (+/- 31,000) assaults (of various kinds), approximately 100,000 (+/- 15,000) cases of deprivation of freedom, and 124,000 (+/- 20,000) cases of war labor violations.⁵

Although physical assault and sexual violations made up the bulk of human rights reporting about violations in Sierra Leone, property destruction was three times more common than assaults, internment, and war labor, and more than ten times more frequent than sexual violations. Given the intense poverty of Sierra Leone, property theft and destruction should be understood as potentially life-threatening. A topic for additional research is the impact of property theft and destruction on mortality: we hypothesize that property destruction and theft is associated with substantially elevated levels of mortality.

2 Historical Background

Sierra Leone's decade of conflict⁶ was characterized by three distinct phases. The TRC identified these phases as 'conventional target warfare' between March 1991-November 1993; 'guerrilla warfare' between November 1993 and March 1997; and 'power struggles and peace efforts' lasting until the end of May 2000. Sierra Leone became completely independent from Great Britain in 1961 (though it remained a member of the Commonwealth), when the country set up its own parliamentary system. By 1969, the party that had led the country to independence, the Sierra Leone Peoples Party (SLPP), had been replaced in power by the repressive All People's Congress (APC), which effectively led a one-party state. Siaka Stevens served as its president until 1985.

Over the years of the APC's control, power and resources gradually became concentrated in the capital, Freetown, while what dissent there was centered on student activism at the University of Sierra Leone. During those years and despite its significant natural resources in the form of diamonds, gold, bauxite, and rutile (titanium ore), Sierra Leone became the poorest country in the world, according to the United Nations (UN).

In the late 1980s, in this climate of poverty, repression, and concentrated state power, the RUF began to establish itself. In next-door Liberia, conflict began in December 1989, when the National Patriotic Front of Liberia (NPFL), led by former president Charles Taylor, launched an attack on the Liberian government. The Liberian civil war, combined with conditions in Sierra Leone itself, helped fuel the RUF, and on March 23, 1991, the RUF, under the leadership of former army corporal and television cameraman Foday Sankoh, invaded the eastern Sierra Leonean district of Kailahun and, simultaneously, the southern district of Pujehun. The TRC estimated that about 80 percent of

⁴The estimation technique is presented in the methodological appendix to this report.

⁵Numbers in text rounded to closest thousand. See Figure 1 for precise estimates and confidence intervals.

⁶A detailed account of the entire conflict is available in the TRC Report, Volume 3A, Chapter Three, pp 89-462.

Figure 1: Estimates of counts of violations during the Sierra Leone armed internal conflict. Source: SLWCD database.

Violation	Weighted Count	95% Confidence Interval
Forced Displacement	1,473,475	1,328,159 - 1,618,791
Destruction of property	253,470	218,318 - 288,623
Property theft	176,939	155,651 - 198,226
Assault & beating	139,283	109,522 - 169,045
Amputation	6,173	2,986 - 9,359
Physical torture	511	36 - 986
Arbitrary detention	95,736	81,389 - 110,083
Disappearance & abduction	2,019	707 - 3,330
Forced recruitment	30,563	24,159 - 36,967
Drugging	17,060	7,976 - 26,144
Forced Labor	75,892	62,751 - 89,033
Rape	24,647	15,949 - 33,345
Sexual slavery	9,660	5,336 - 13,984
Sexual abuse	5,030	1,767 - 8,293
Forced consumption	14,473	6,626 - 22,320
Overall Total	2,324,931	2,092,579 - 2,557,283

the RUF's initial force of 2,000 belonged to the NPFL or were Burkinabes (citizens of Burkina Faso).⁷

This invasion began the first of the three phases of the conflict: 'Conventional Target Warfare', March 23, 1991-November 13, 1993, which ended with the RUF's loss of the border town of Baidu in Kailahun district, the apparent defeat of the National Provisional Ruling Council (NPRC), and Sankoh's memo announcing reversion to 'jungle warfare.' As can be seen in Figure 2, violence (as measured by total reported killings)⁸ began at a high level but declined toward 1993. In the succeeding months of 1991 and early 1992, the RUF pushed forward into the diamond mining district of Kono. The SLA began a recruitment of volunteers that enlisted predominantly urban youths from Freetown; army units stationed at the frontier also recruited local volunteers (vigilantes, border guards, and others).⁹

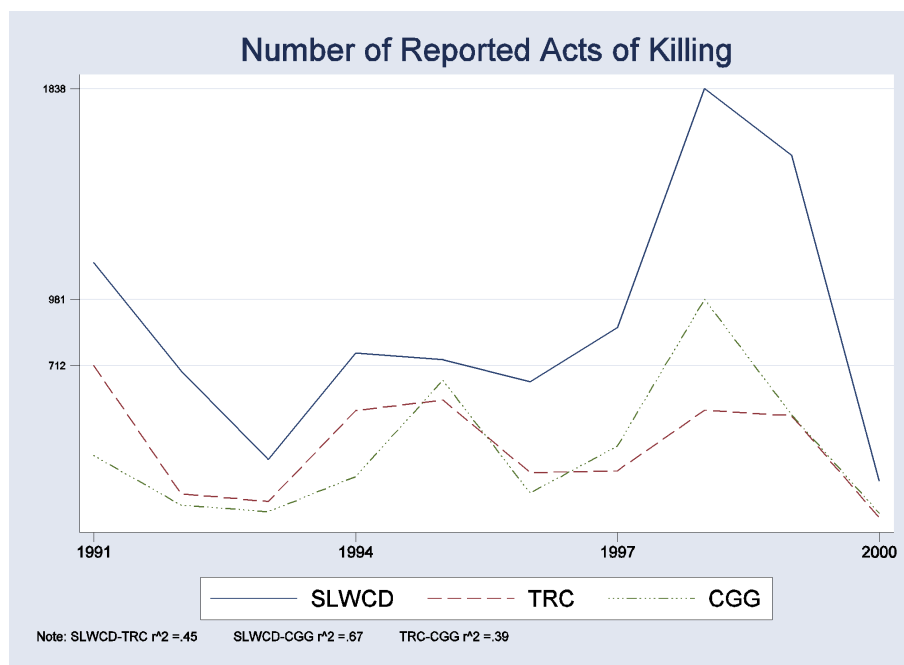
On April 29, 1992, a group of junior military officers from the Sierra Leone Army (SLA) staged a coup, replacing the elected government with its own creation, the NPRC. Stevens' successor as head of the APC, Major General

⁷TRC Report, Vol 3A, p119, paragraph 126.

⁸While the total reported killings presented in Figure 2 do not represent estimates of total killing, and may be subject to bias, the fact that the data from all three data sources track so closely is compelling evidence that the true pattern of killings over time is similar to the one presented.

⁹TRC Report, Vol 3A, p199, paragraph 500.

Figure 2: Counts of reported killings during the Sierra Leone armed internal conflict. Source: SLWCD, TRC, and CGG.



Joseph Saidu Momoh, was sent into exile in Guinea and replaced by twenty-five-year-old Captain Valentine Strasser as head of state. In December 1992, the NPRC convicted SLA members of treason and executed them.

The SLA fought back throughout 1993, driving the RUF back to the borders, and Sankoh circulated a memo announcing a new strategy of guerrilla tactics. At the end of 1993, Strasser declared a ceasefire, ending the first phase of the conflict and suspending the suffering for civilians, temporarily. Figure 3 shows the total estimated people who were displaced from their homes in each year of the conflict, with the shaded area indicating the margin of error around each annual estimate. It also shows that the calm of 1993 was short-lived. Joining the expanded SLA in the fight against the RUF were Civil Defense Units (CDUs), formed between 1991 and 1993 as traditional authorities in local areas appointed well-known and experienced people to form local militias. The second phase of the conflict began in November 1993 as the war entered a guerrilla mode. In 1996 the Kamajor Society, a local militia (or ‘warrior society’¹⁰) previously concentrated in the Bonthe District and primarily made up of Mende, emerged onto the national stage, generally fighting on the side of the RUF. This group broadened its makeup by initiating civilians after a secret ceremony that involved a ‘rigorous series of physical and psychological challenges’¹¹ using organs, tissue, blood, and flesh from the bodies of dead people; this ceremony was believed by initiates to confer supernatural power.¹² At the beginning, initiates were nominated by chiefdom authorities and sent back to their area of origin after initiation. However, the rate of initiation increased rapidly throughout 1995 and 1996 until almost every chiefdom had its own group of Kamajors.¹³ The CDF that emerged in late 1995/early 1996 began as a union of all these local militias, including

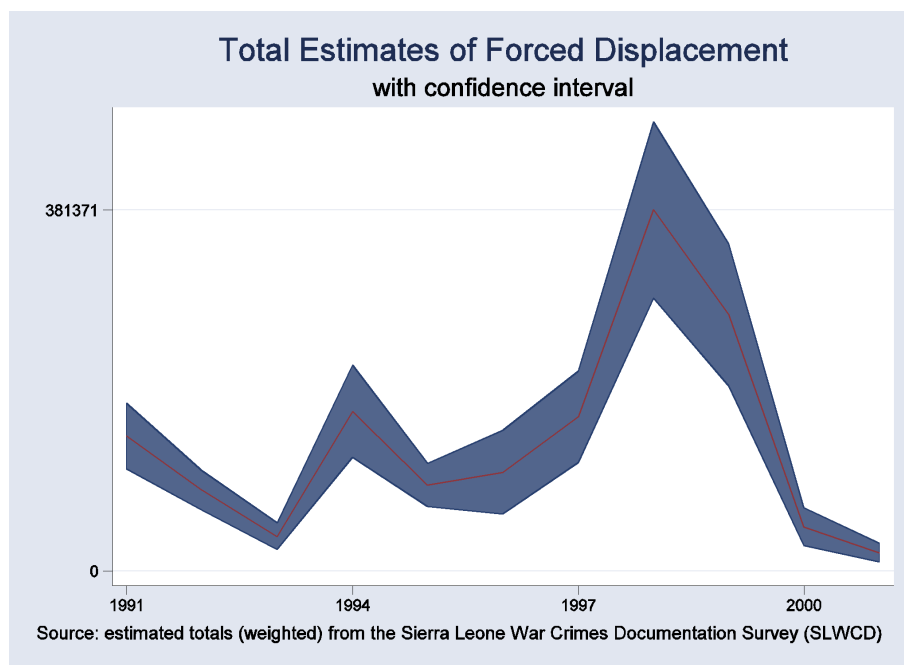
¹⁰TRC Report, Vol 3A, p214, paragraph 562.

¹¹TRC Report, Vol 3A, p495, paragraph 118.

¹²TRC Report, Vol 3A, p215, paragraph 566.

¹³TRC Report, Vol 3A.

Figure 3: estimates of Forced Displacement during the Sierra Leone armed internal conflict, with 95% confidence interval



the Kamajors, who became its main force in fighting the RUF. Over time, rivalry developed between the SLA and CDF.

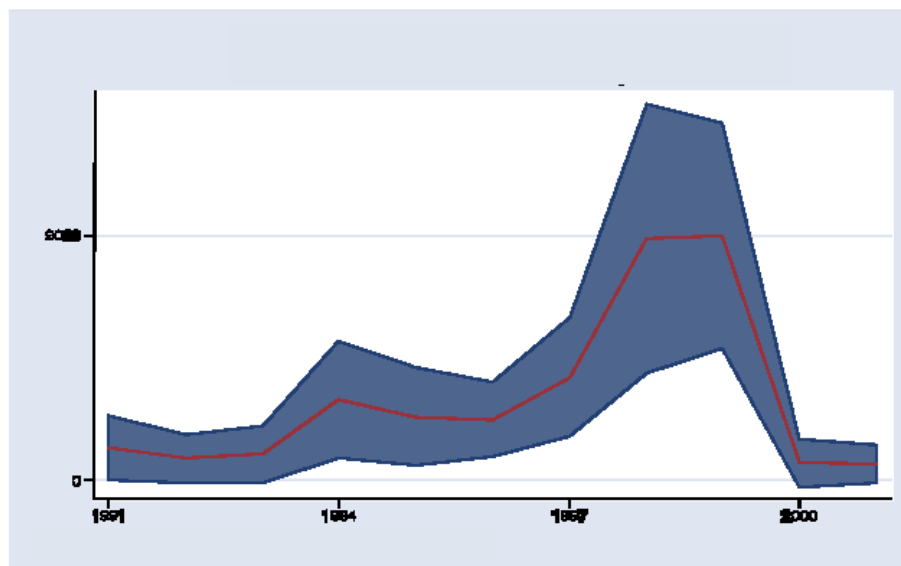
In addition to the above actors, both the NPRC and Ahmad Rejan Kabbah's exiled government hired private military companies mercenaries for short periods to assist their efforts during this period.¹⁴ Most of the parties also engaged in forced recruitment of men of all ages (see the age analysis below) to support their campaigns. shows that the substantial increase in the use of forced recruitment was in 1994, but it was to be a feature of the conflict thereafter. The conflict spread across the country, nearly reaching Freetown by 1995, by which time the RUF had expanded its coverage so broadly that it had established a presence in each of Sierra Leone's 12 districts.¹⁵ The NPRC then drove the RUF back, and agreed to hand over power to a democratically chosen civilian government. Kabbah, a former UN diplomat, won the presidential election in April 1996, and the Sierra Leone Peoples Party (SLPP) won a majority in the parliamentary election. Over the ensuing months, the SLPP managed to negotiate the Abidjan Peace Accord with the RUF, which was signed on November 30, 1996, by both Kabbah and Sankoh.¹⁶ After disagreements over disarmament and the creation of a monitoring force, however, the 1996 Kabbah-Sankoh agreement quickly broke down. On May 25, 1997, SLA officers staged a coup, forming the AFRC, led by Major Johnny Paul Koroma and overthrowing President Kabbah, who escaped to Guinea by helicopter. The AFRC, which immediately suspended the Constitution of Sierra Leone, had the support of most of the SLA, though some remained loyal to the elected government. The RUF leadership joined the government at the AFRC's invitation. Koroma was sworn in as head of state on June 17, 1997. The CDF was ordered to disband following the May 1997 military coup but continued fighting the RUF/AFRC.

¹⁴TRC Report, Vol 3A, p203, paragraph 291.

¹⁵TRC Report, Vol 2, p42, paragraph 137.

¹⁶'Peace Agreement between the Government of the Republic of Sierra Leone and the Revolutionary United Front of Sierra Leone (RUF/SL).'

Figure 4: Estimates of Forced Recruitment during the Sierra Leone armed internal conflict, with 95% confidence interval. Source: SLWCD.



The second phase of the conflict ended in March 1997, when Foday Sankoh was taken into custody by Nigerian forces. The third phase of the conflict, which lasted until May 2000 when the RUF was stripped of its presence in government, saw some of the highest levels of violence, particularly of displacement, theft and destruction of property. The notable increase in property destruction in Phase III of the conflict is shown in Figure 5. In February 1998, after an escalating military standoff and another failed peace plan, the RUF/AFRC government was overthrown. This time, the intervention came from the non-standing military force Economic Community of West African States Military Observer Group (ECOMOG). The ECOMOG was originally created in 1990 as a peacekeeping force to intervene after the state structure in Liberia collapsed. It was formed from units of national military forces in member states of the Economic Community of West African States (ECOWAS), and in 1992 a contingent of the SLA made up part of the ECOMOG force in Liberia.

ECOMOG first became active in Sierra Leone in June 1997, shortly before ECOWAS, supported by the UN Security Council, imposed a trade, arms, and petroleum products embargo on Sierra Leone.¹⁷ ECOMOG was given the mandate of monitoring and supervising ceasefire obligations and enforcing the sanctions and embargo against the RUF/AFRC regime. On October 21, the parties agreed to a ceasefire, and on October 23, 1997 the six-month ECOWAS peace plan, which mandated the return of the elected government by April 22, 1998, was signed in Conakry, Guinea.¹⁸

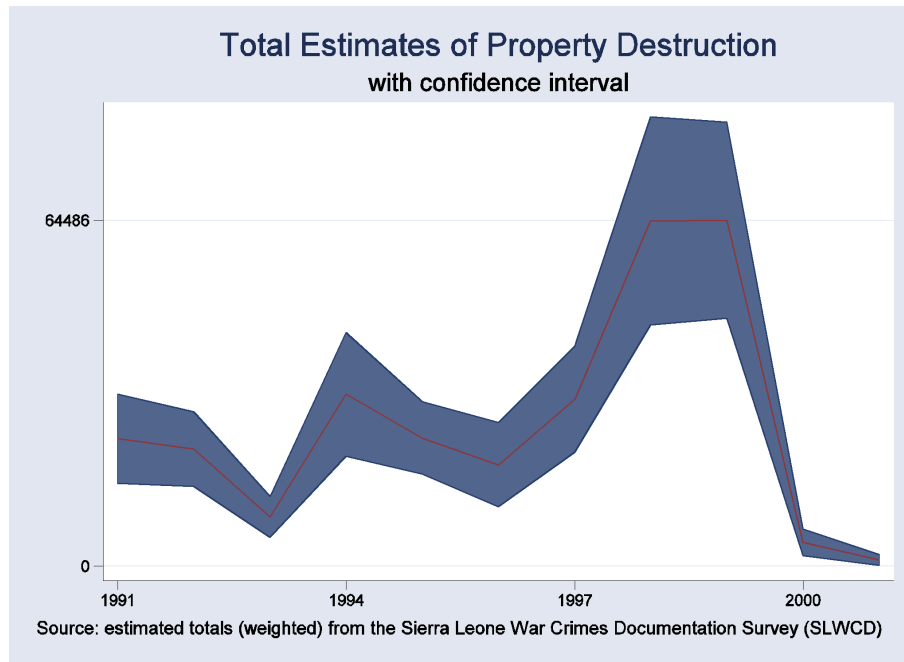
President Kabbah was reinstated in March 1998, and most of the AFRC members left Freetown for the provinces; others surrendered to ECOMOG, and many of the senior officers were charged with treason, convicted, and executed in 1998. The CDF was placed under the control of ECOMOG. During this period, the Commission reported attacks against both other civilians and AFRC supporters by civilians seeking revenge.¹⁹ Two months after his reinstatement,

¹⁷Resolution 1132 (1997) adopted by the UN Security Council at its 3822nd meeting, October 8, 1997.

¹⁸'ECOWAS Six-Month Peace Plan for Sierra Leone 23 October 1997 22 April 1998 (Schedule of Implementation)'.

¹⁹TRC Report, Vol 3A, p304, paragraph 922ff.

Figure 5: Estimates of property destruction during the Sierra Leone armed internal conflict, with 95% confidence interval. Source: SLWCD



President Kabbah put the former ECOMOG commander in charge of creating a new army. The old SLA was disbanded in July 1998, and the new SLA began fighting alongside ECOMOG against the RUF/AFRC forces. In September 1998 the president restored the CDF's authority by directing the appointment of CDF administrators in every district to liaise with the SLA.

On January 6, 1999, the RUF/AFRC forces invaded Freetown in an attempt to regain control. ECOMOG and the CDF fought back, and after an increasing rift between the RUF and AFRC, on July 7, 1999 President Kabbah and RUF leader Foday Sankoh (who had been convicted of treason in Freetown in 1998 but was freed after the January 1999 attack on Freetown) signed the Lomé Peace Accord, which made Sankoh Vice President and Chairman of a Commission for the Management of Strategic Resources, National Reconstruction and Development, and gave other senior RUF members positions in the government. Koroma did not take part in negotiating the peace agreement, but he was made chairman of a government body, the Commission for the Consolidation of Peace.

Nonetheless, his absence from the negotiations instigated the formation of a splinter group from the RUF/AFRC alliance, the West Side Boys, who were active between October 1998 and the end of 2000. This group preyed on civilians in Freetown²⁰, took officials from ECOMOG, NGO, and the United Nations Observer Mission in Sierra Leone (UNOMSIL) hostage in 1999, and also kidnapped 11 UK Royal Marines in late August of that year. The resulting British military response largely destroyed them. Although the West Side Boys claimed Koroma as their figurehead, he publicly disassociated himself from them and officially disbanded the AFRC in August 2000.

²⁰TRC Report, Vol 3A, p330, paragraph 1036.

The Lom Peace Accord called for an international peacekeeping force to be jointly provided by ECOMOG and the UN, which established the United Nations Mission in Sierra Leone (UNAMSIL), withdrawing its predecessor, UNOMSIL, which had been established in June 1998.²¹ ECOMOG withdrew in 2000 and was gradually replaced by UNAMSIL forces. By then, the RUF had begun to violate the agreement, eventually taking nearly 500 UNAMSIL personnel hostage and capturing their arms and ammunition, and there was infighting between the RUF and AFRC. On May 8, 2000, members of the RUF shot and killed as many as 20 people demonstrating against the RUF violations outside Sankoh's house in Freetown. Koroma called on his forces to support the government against the RUF, and in May and June 2000 the AFRC fought with the SLA against the RUF, while the UN Security Council increased UNAMSIL's presence to 13,000. Sankoh and other senior members of the RUF were arrested and the group was stripped of its positions in government.

A new ceasefire was signed in Abuja in November 2000 and, after fighting continued, a second in May 2001. Disarmament proceeded slowly, and President Kabbah declared the civil war finally and officially over on January 18, 2002.

In March 2003, Sankoh was indicted in a Special Court for Sierra Leone on 17 counts of war crimes (controversially, the indictment covered only events after 1996). He died of a stroke in July 2003 while awaiting trial. Unfortunately, Sierra Leone is far from recovered from the decades of corruption prior to and the violence during its armed internal conflict. The UN's 2005 Human Development Report notes Sierra Leone's continued poverty, ranking it 170th of 177 countries, while it comes in second to last out of 177 countries on the UN Human Development Index 2005, which is a composite measure of GDP, literacy and life expectancy.²² Average life expectancy is 39.87 years,²³ up from 34 in 1970-75, and in 2003 approximately a quarter of all children did not live to see their fifth birthdays.²⁴

3 Phases of the Conflict and the RUF

Figure 2 shows the pattern of all killings reported over the years from 1991 to 2000 for each of three datasets, the collection of testimony gathered by the TRC and that collected by CGG, and the SLWCD survey.²⁵ In all three datasets, three peaks of violence are clearly visible in 1991, 1994-1995, and 1998-1999, matching the Commission's phase structure. All three datasets show the same broad peaks of violence at the same times but to varying degrees. It is particularly interesting that the peak structure in the CGG and SLWCD data both support the Commission's hypothesis even more strongly than the Commission's own dataset does. This analysis also supports the Commission's claim that the violence was most intense during Phases I and III of the conflict.

The Commission's own data does not correlate as closely with either of the other two datasets as they do with each other. The pattern is similar, but not as pronounced. As has been the case in previous studies,²⁶ additional data modifies earlier findings based on the first set of data. It is often the case that additional data clarifies a pattern which was

²¹ UN Security Council resolution

²² Human Development Report 2005, p227.

²³ CIA World Factbook.

²⁴ UNDP.

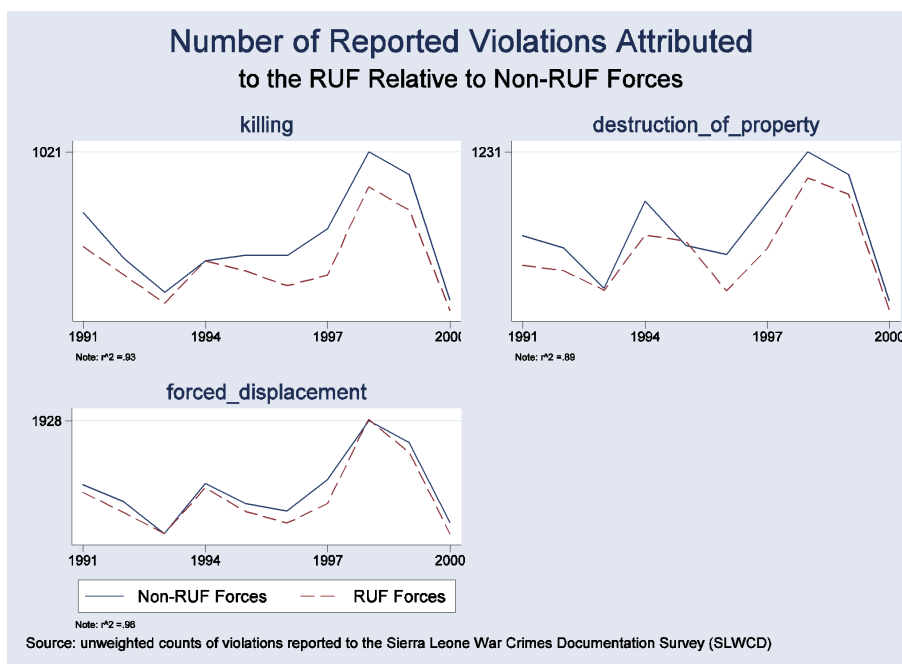
²⁵ This graph shows reported killings, not the estimated totals.

²⁶ See for example Romesh Silva and Patrick Ball, *The Profile of Human Rights Violations in Timor-Leste, 1974-1999*, The Commission for Historical Clarification in Guatemala *Guatemala: Memoria de Silencio*; and Patrick Ball et al., *Killings and Refugee Flow in Kosovo March- June 1999*.

obscure when observed with less data.

Figure 6 compares the number of reported violations attributable specifically to the RUF versus those attributed to non-RUF forces throughout the years of the conflict for killings, destruction of property, and forced displacement as reported to the SLWCD. All three violation types – killings, destruction of property, and forced displacement – show the same broad peaks as total killings in Figure 2, and in all three violation types, the RUF had nearly as many reported violations as all the other perpetrators combined.²⁷ The Commission found the RUF to have been responsible for the

Figure 6: Counts of reported violations attributed to the RUF relative to those non attributed to the RUF during the Sierra Leone armed internal conflict. Source: SLWCD.



largest number of human rights violations in the conflict, both for the period 1991 to 2000 as a whole and for each year individually,²⁸ as well as the majority of all violations in every single one of Sierra Leone's 12 districts during Phase II.²⁹

In general, the Commission also found that the reported violations attributable to the different actors were not evenly spread across time, as might be expected given the actors' movements in and out of power.³⁰ The reported violations attributed to the AFRC, for example, were concentrated in the years 1997 to 1999, the years it was in power, while the number of reported violations attributed to the CDF also peaked in 1997 to 1999. Similarly, no violations were attributed to the SLA during the years 1998 to 1999, when it was disbanded and recreated, but more violations were attributed to the SLA in the first phase of the conflict. At that time, the SLA was being rapidly expanded. No Peace without Justice (NPWJ), an international NGO, claims, for example, that the NPRC's intensive recruitment to the SLA

²⁷This graph shows reported violations, not the estimated counts of violations.

²⁸TRC Report, Vol 2, p28 paragraph 29.

²⁹TRC Report, Vol 2, p42, paragraph 133.

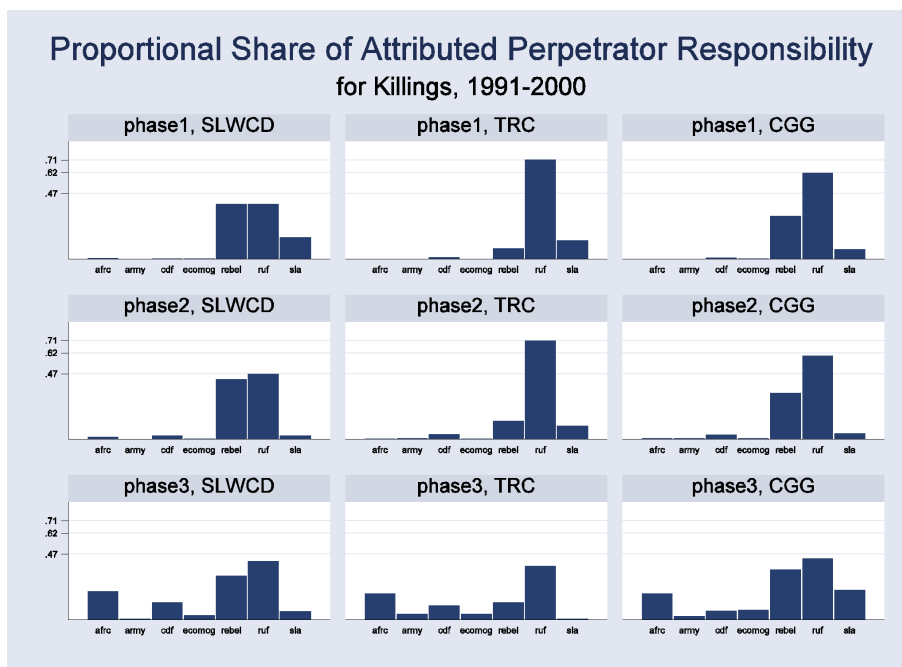
³⁰TRC Report, Vol 2, p29.

during 1993 led to hastily and poorly trained soldiers who were unable to contain the advance of the RUF or protect civilians from the RUF's attacks. 'Many called the SLA 'So-bels', or 'Soldier-Rebels', noting that in some cases there was little to distinguish the two.'³¹

In Phase III, by contrast, TRC identified many more actors who committed violations.³² As documented by the Commission, the largest proportion of reported killings are always attributed to the RUF, and the additional data support the pattern found by the Commission. However, in addition to the 60.5 percent (24,353/40,242) of reported violations the Commission found were attributable to the RUF, it attributed 9.7 percent³³ of violations to 'Rebels', as shown above, and notes that, 'Typically the term describes RUF fighters and ex-SLA fighters loyal to the AFRC'.³⁴ This may be true, but it is difficult to confirm with the existing data. The Commission also noted occasions when the RUF carried out attacks dressed in full SLA military uniform and succeeded in deceiving the local population into believing the SLA was responsible for the RUF's attacks.³⁵ The extent to which this affected interviewees' beliefs about the perpetrators of their violations is also difficult to confirm.

Nevertheless, the new data are consistent with the claim that the RUF determined the course of the conflict in time and space; even though the RUF was not responsible for as much as 80 percent of the reported killings, it was responsible for the overall pattern. Our analysis suggests that approximately 75 percent of total violations were committed by the RUF and the Rebels.

Figure 7: Proportional Share of Attributed Perpetrator Responsibility, based on unweighted counts, during the Sierra Leone armed internal conflict. Source: SLWCD, TRC and CGG.



³¹NPWJ, Conflict Mapping, p36.

³²TRC Report, Vol 3A, p552-3, paragraphs 330-336.

³³TRC Report, Vol 2, p 38, footnote 22. No detailed numbers are given.

³⁴TRC Report, Appendix 1, p8.

³⁵TRC Report, Vol 2, p43, paragraph 138.

4 Geography and Violence

Article 7 of the Rome Statute of the International Criminal Court defines a ‘crime against humanity’ as any of a number of acts – murder, enslavement, extermination, forcible transfer of population, rape, sexual slavery – when they are committed as ‘part of a widespread or systematic attack directed against any civilian population, with knowledge of the attack.’³⁶ The terms ‘widespread’ and ‘systematic’ have been applied in a number of cases, most notably in the International Criminal Tribunal for Rwanda (ICTR) and the International Tribunal for the Former Yugoslavia. The ICTR states, for example, ‘A widespread attack is an attack on a large scale directed against a multiplicity of victims, whereas a systematic attack is one carried out pursuant to a preconceived policy or plan’.³⁷

In its report, the Commission claimed that reported killings were geographically concentrated in 1991 to 1993 and became more widespread thereafter. In the years when violence was more severe – 1991, 1994-1995, and 1998-99 (as previously seen in) – it is clear that reported killings were more concentrated in a few regions than in the intermediate years.

Sierra Leone is divided into four provinces: Northern Province, Southern Province, Southern Province, and Western Area (which includes only the country’s capital, Freetown, and its environs). The first three of these provinces are further divided into 12 districts, which are in turn subdivided into 149 chiefdoms. Figure 8 presents the number of reported killings across the four regions. It shows that during the first phase of the conflict (between 1991 and 1993) reported killings were concentrated in a few regions. However, during Phase II, violence (as measured by reported acts of killing) spread to all regions of Sierra Leone but at a substantially lower level than during Phase I. While this finding is consistent with that of the Commission that as the conflict progressed, it spread to more regions of the country, this observation complicates the argument that violence was widespread. During Phase III, reported killings increase again, but concentrated in specific regions each year: in the East and North during 1998, and then in the North and West during 1999. The concentration of reported killings in the West during 1999 may show the effect of the invasion of Freetown by the RUF in February.

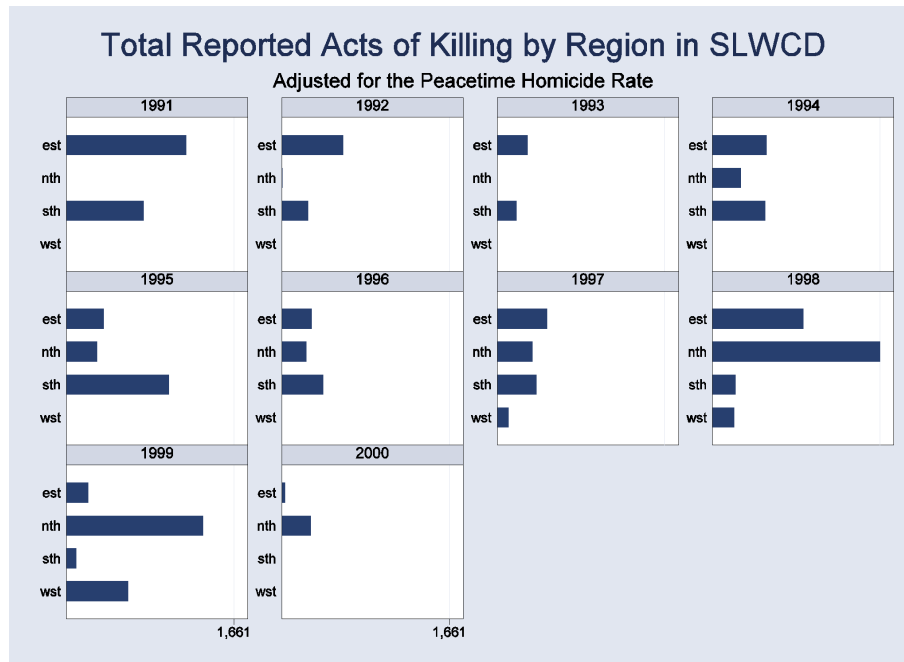
When reported killings are at their peak, they are concentrated in certain regions during certain episodes of the conflict; they are not uniformly distributed across all regions and all times. In particular, reported killings are concentrated in 1991-1992 in the East and South, in 1995 in the South, and in 1998-1999 in the North and West. In 1999 in particular, almost all the killings reported to CGG and TRC took place in Freetown. Recall that the SLWCD Survey is obtained from a randomly selected sample, which can be more representative than a convenience sample composed of people who choose to give their statements (such as the TRC and CGG samples). Figure 8 shows that in 1999, more killings were reported in the North than in Freetown. The difference may be due to inequality of access: the convenience samples may have had greater access to people in Freetown than those outside it.

³⁶Rome Statute of the International Criminal Court.

Web: http://www.un.org/law/icc/statute/99_corr/2.htm (2006-03-06).

³⁷Judgement and Sentence in case ICTR-95-1A-T, Ignace Bagilishema.

Figure 8: Reported non-combatant killings by region during the Sierra Leone armed internal conflict, adjusted for peacetime homicide rate. Source: SLWCD.



5 Perpetrators

The Commission found that the most common reported violations were forced displacement, abductions, arbitrary detentions, and killings.³⁸ The new data available from the SLWCD further explain the peak the Commission found at the end of the conflict, when violations were reported to have been perpetrated by actors other than the RUF. The following six graphs, Figure 9 to Figure 14, show raw counts of reported violations by perpetrator for killings, forced displacement, and destruction of property by the major perpetrator groups. The first set of graphs in each pair of figures (that is, Figure 9, Figure 11, and Figure 13) shows the pattern of reported violations attributed to each of the six main actor groups; the second set of graphs for each violation type (that is, Figure 10, Figure 12: Reported acts of forced displacement attributed to particular perpetrator groups during the Sierra Leone armed internal conflict, in standardized scale. and Figure 14) shows the same patterns but with the vertical scale of each graph standardized to show more clearly the relative proportion of responsibility attributed to different perpetrators. For each violation type – killings, forced displacement, and destruction of property, three notable patterns can be observed. First, of all the main perpetrator organizations, the RUF is the most responsible for these reported violations in all phases of the conflict. Second, the only group to which similar (though still lesser) levels of reported killings, forced displacement, and property destruction in the three main phases of the conflict is attributed is ‘Rebels’. Third, while reported violations attributed to the RUF and to Rebels are spread across three waves (corresponding to the phases of the conflict as defined by the Commission), almost all of the reported violations attributed to the CDF, AFRC, and other groups are concentrated in the final phase of the conflict. Reported killings attributed to the SLA and other institutions are reported in Phases I and III of the conflict. This analysis, too, is consistent with the view that the RUF drove the

³⁸TRC Report, Vol 2, p28, paragraph 25.

Figure 9: Reported acts of killing attributed to particular perpetrator groups during the Sierra Leone armed internal conflict. Source:SLWCD.

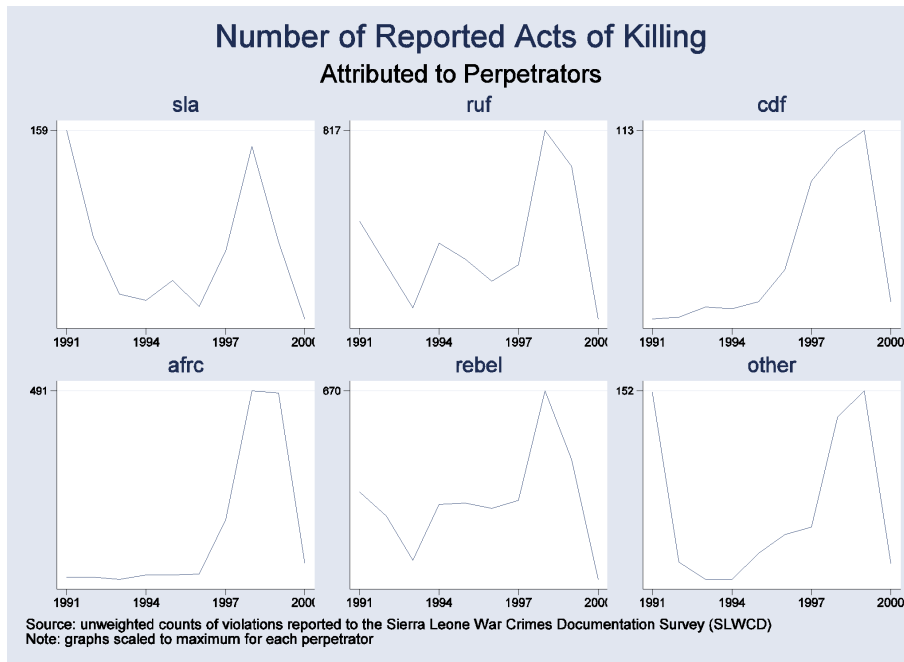


Figure 10: Reported acts of killing attributed to particular perpetrator groups during the Sierra Leone armed internal conflict, in standardized scalet. Source: SLWCD.

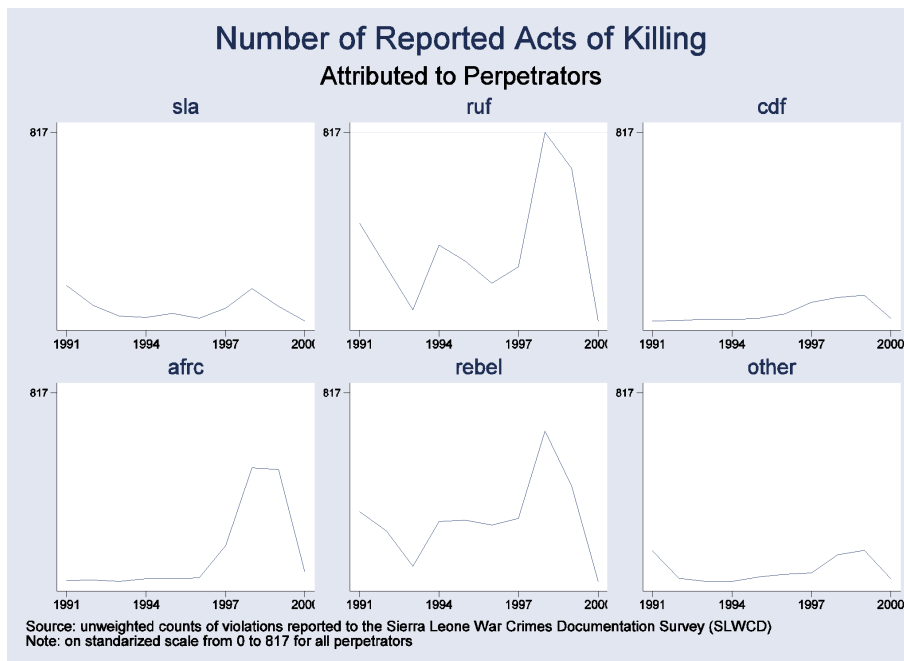


Figure 11: Reported acts of forced displacement attributed to particular perpetrator groups during the Sierra Leone armed internal conflict.

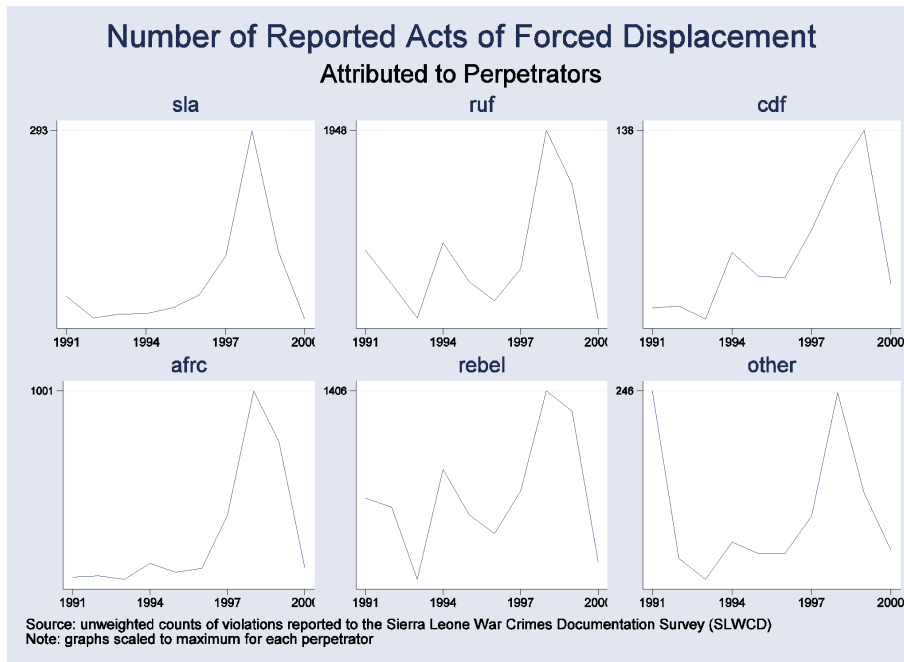


Figure 12: Reported acts of forced displacement attributed to particular perpetrator groups during the Sierra Leone armed internal conflict, in standardized scale.

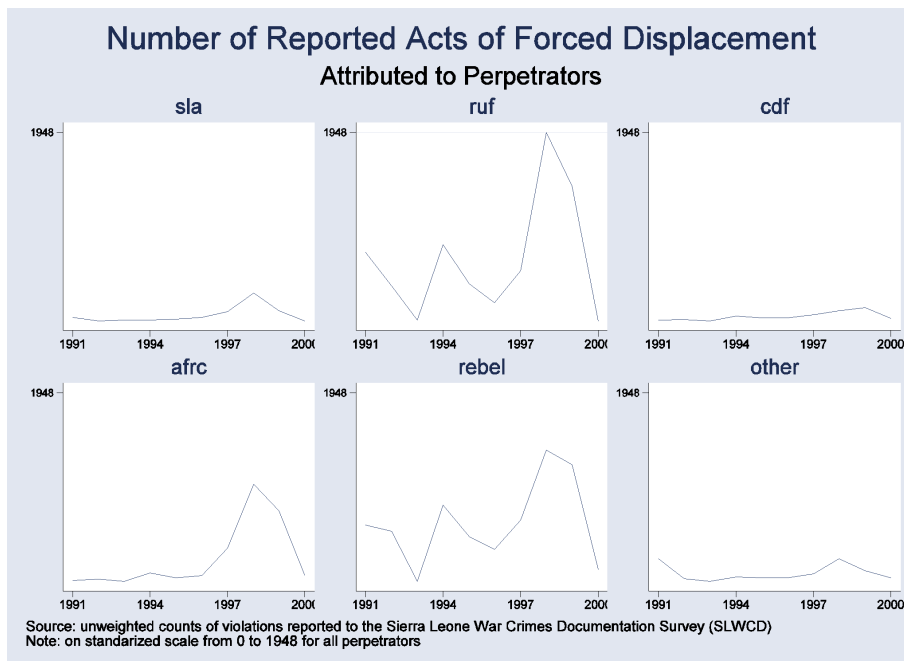


Figure 13: Reported acts of property destruction attributed to particular perpetrator groups during the Sierra Leone armed internal conflict, in standardized scale.

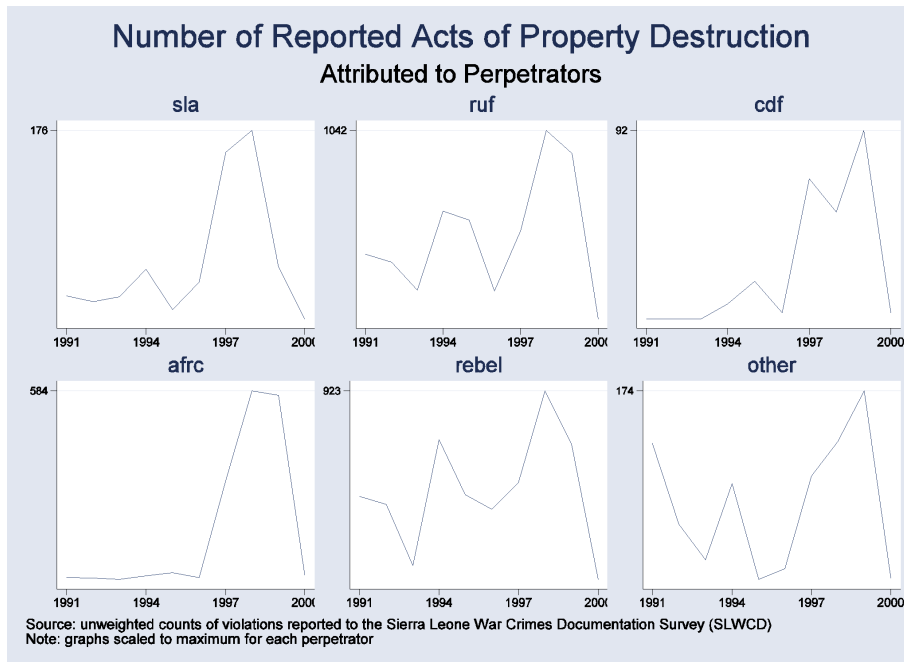
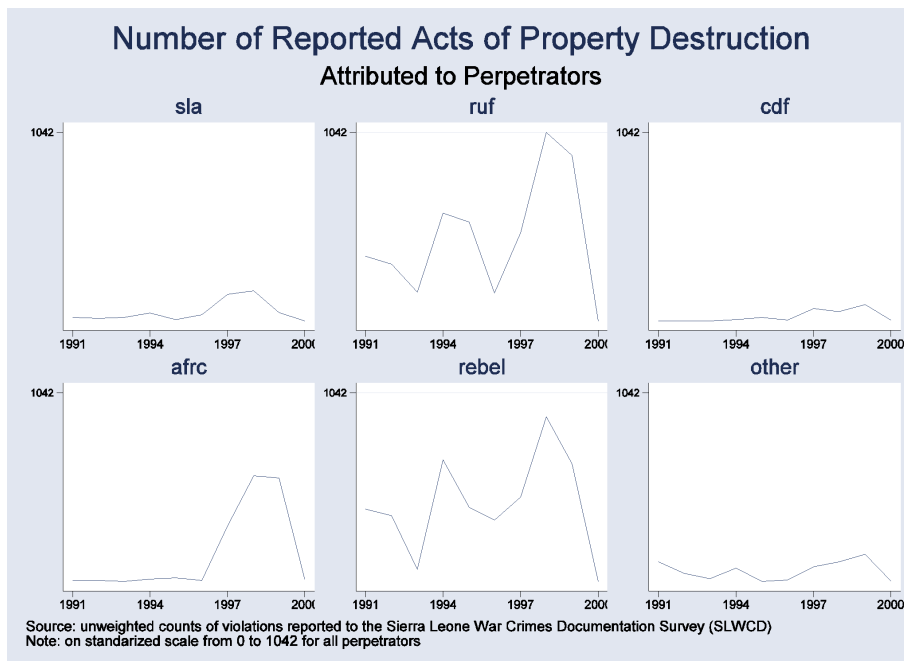


Figure 14: Reported acts of property destruction attributed to particular perpetrator groups during the Sierra Leone armed internal conflict, in standardized scale.



nature of the conflict. Even though the RUF was not responsible for all of the violations all of the time, Figure 10, Figure 12, and Figure 14 show that, as the Commission found, the RUF is responsible for more reported killings, displacements, and acts of property destruction than any other group at all times. Other perpetrator groups become active only in Phases II and III of the conflict, and they never approach the RUF in terms of the magnitude of their violence.

6 Targeting by Ethnicity

In its report, the Commission seemed to contradict itself on the question of whether specific groups were targets of violence. It said both that civilians were ‘express targets of militias and armed groups’³⁹ and that killings were indiscriminate;⁴⁰ targeting and indiscrimination would be opposite strategies. The Commission claimed that the following groups were targeted during the conflict: heads of household; the elderly; those who were affluent or possessed of elite status; those who were the targets of grudges or vendettas; Nigerians; in the south, people from the north.⁴¹ However, most of these findings were not substantiated by the Commission’s quantitative data.

Sierra Leone’s population is home to as many as 20 native African tribes, of whom the two largest are the Temne (approximately 30 percent of the population) and the Mende (also approximately 30 percent). About 10 percent are Krio (descendants of freed Jamaican slaves who were settled in the Freetown area in the late 18th century). A mix of other tribes and groups from Guinea and Liberia, and a small number of Europeans, Lebanese, Pakistanis, and Indians makes up the remaining 30 percent.⁴² However, these ethnic groups are not evenly distributed across the country. The Mende, for example, are concentrated in the south, and the Temne in the north. A claim that a particular ethnic group was disproportionately targeted must accordingly take into account the uneven distribution of these various ethnic groups.

Households in the SLWCD were selected to represent the various areas of Sierra Leone, and it is therefore possible to make conclusions about the ethnic distribution of killings during the conflict across the country by region and district.⁴³ For each district, Figure 15 shows the ratio of the estimated proportion of the population which is of each ethnicity compared against the proportion of victims of killing that are of that ethnicity. The red line in each graph indicates parity – the point at which the proportion of killings equals the proportion in the population. Bars shorter than the red line indicate groups that suffered less than the proportion that would be expected from their population level, and bars longer than the red line indicate ethnicities that suffered higher rates of killing than would be expected. The SLWCD data suggests that in fact killings were not disproportionately targeted at any one group throughout Sierra Leone. Rather, in certain districts particular ethnic groups were killed at disproportionate rates compared to others. For example, a disproportionate amount of killings was reported against a few ethnic groups in specific districts: the Limbas in Kambia, the Madingo in Koinadagu, and the Krio in Freetown. Fewer killings were reported than would

³⁹TRC Report, Vol 3A, p550, paragraph 313.

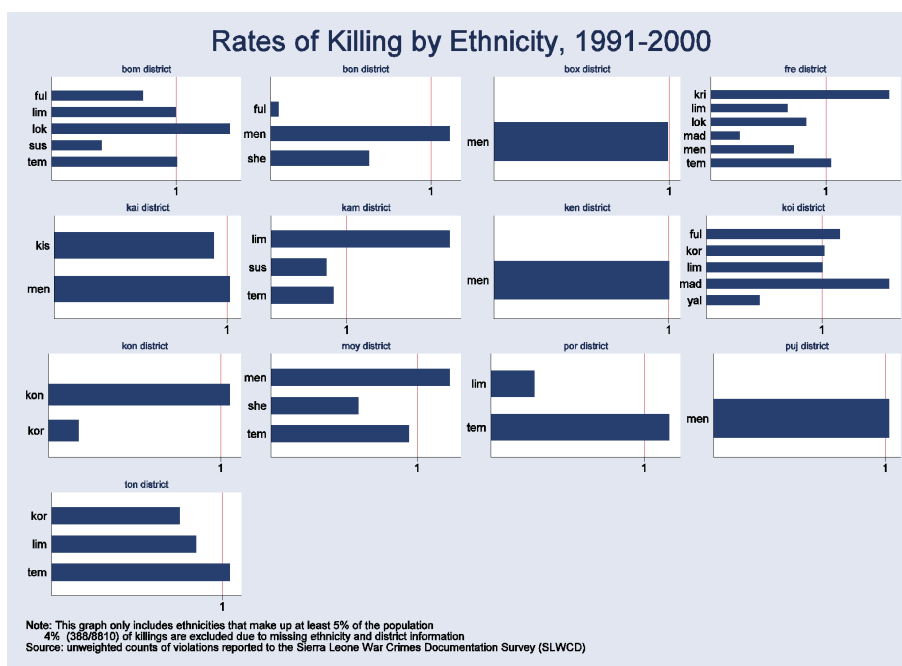
⁴⁰TRC Report, Vol 3A, p554, paragraph 340.

⁴¹TRC Report, Vol 2, p11, paragraph 32.

⁴²CIA World Factbook.

⁴³Each of the conclusions in this and subsequent sections were evaluated using both unweighted (raw) counts from the survey and the weighted projections. This test takes into consideration the uneven weighting among respondents in different areas that reflects the varying number of respondents and the varying number of people in each area. None of the weighted projections contradicted the conclusions drawn from the raw counts. In order to maintain comparability with the TRC and CGG data, the SLWCD unweighted counts are used in the analysis.

Figure 15: Source: Counts of violations reported to the SLWCD.



have been proportionately expected against two groups: the Susu in Bombali and the Kor in Kono.⁴⁴

The Commission found that in the south ethnic targeting was carried out against people from the north; the Commission noted that it began with targeting soldiers, since by the beginning of the war most of the members of the armed forces were from the north.⁴⁵ The Commission's findings are, however, inconsistent in several respects with those in Figure 15, which show no targeting of northerners in the south. The Limbas in Kambia were northerners in the north. Further study may help determine whether the Limbas were targeted by one particular perpetrator whose actions against this group expanded out of the south. However, the SLWCD data suggests that ethnic targeting was of a more localized nature than was previously understood.

These data also help support at least partially another of the Commission's claims, that the RUF targeted Lebanese, Fullahs, Madingos, and Marrakas: 'These groups are essentially trading or mercantile communities who were targeted because of their perceived wealth and the opportunity to appropriate their property.'⁴⁶ However, the Commission itself did not have sufficient quantitative data to support this hypothesis. The Commission also does not discuss the Krios in its list of groups targeted by the RUF. However, the Krios are the majority and the middle class in Freetown, and may also have been seen as wealthy enough to be worth targeting for theft, like the trading or mercantile groups the Commission did discuss.

⁴⁴ Because up-to-date ethnic census data are not available, the error rate associated with this finding is unknown and we can only be confident about the most extreme patterns.

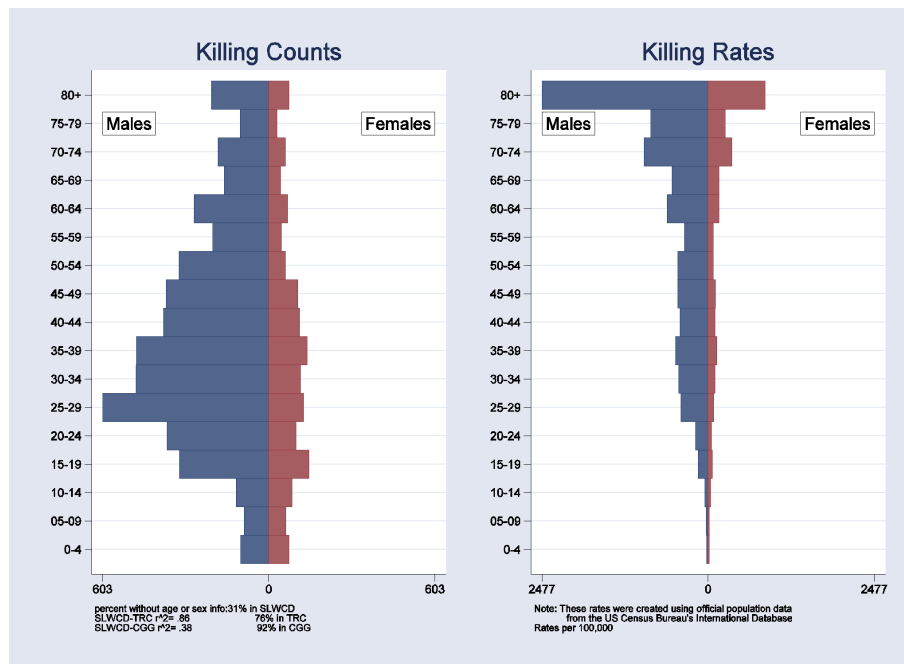
⁴⁵ TRC Report, Vol 3A, p519, paragraph 202.

⁴⁶ TRC Report, Vol 3A, p523, paragraph 218.

7 Targeting by Age and Sex

Figure 16 compares the raw counts from SLWCD of reported killings by age and sex to the proportion of the general population those counts represent. While the largest number of reported killings were males aged between 25 and 29, proportionately the group reported to have lost the largest percentage of its number was older men, who form a small minority of Sierra Leone’s population. As of 2005, the median age for males in Sierra Leone was 17.2, and 44.7 percent of the population is under 15. Only 3.3 percent of the population is over 65.⁴⁷ Because the Sierra Leone population is so young, the age-specific risk for violations can be very different from the age-counts for that violation. Another useful observation about these graphs is that the age and sex distribution found for the SLWCD closely ap-

Figure 16



proximates that found by the TRC. As a result of the probing techniques taught to the interviewers (see the Appendix), the SLWCD had very little missing data on the age and sex of the victims, whereas the TRC was missing the age and sex for half or more of the victims. The fact that the SLWCD correlates with the TRC with an r² of 0.68 indicates that the TRC’s missing data did not excessively distort its results.

7.1 Sexual Slavery, Sexual Abuse

Three types of violations in Sierra Leone have gotten the most attention worldwide: amputations, sexual slavery, and sexual abuse. The Commission found that ‘[t]he RUF carried out widespread rapes and acts of sexual violence against women and girls.’⁴⁸ And that, in Phase II: ‘The RUF carried out widespread rapes and acts of sexual violence in every

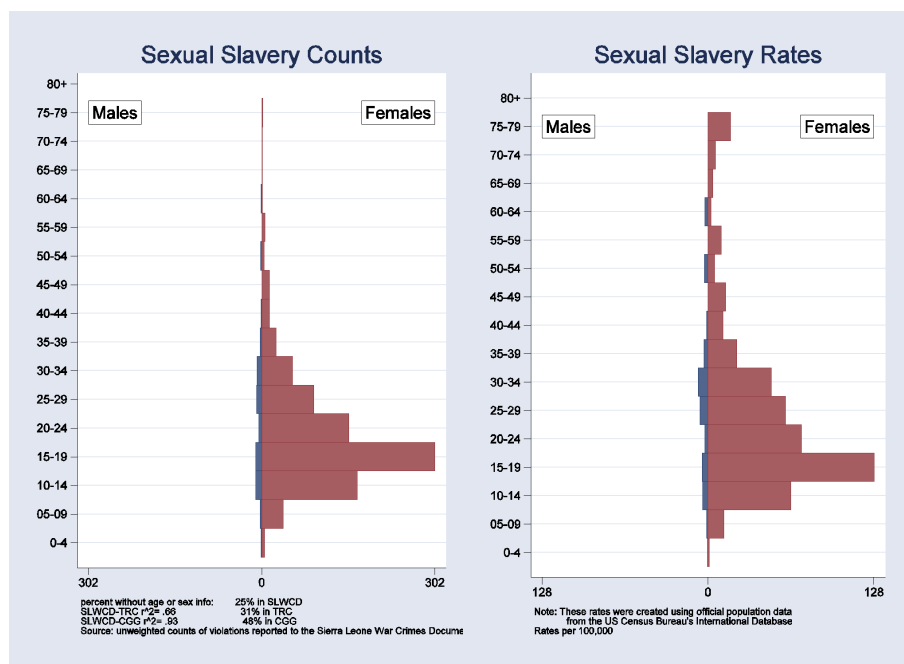
⁴⁷CIA World Factbook.

⁴⁸TRC Report, Vol 2, p40, paragraph 118.

community it entered.’⁴⁹ Human Rights Watch, in 1998, also claimed, ‘sexual violence has been widespread, against thousands of women and girls.’⁵⁰

Sexual slavery and rape are separate phenomena rape is typically a one-time occurrence, while sexual slavery is a series of acts over a continuous time frame but they seem to be driven by the same policy or practice, as evidenced by the same age and sex distributions, which appear below. The Commission found that girls between the ages of ten and 14 were targeted for rape and for sexual slavery.⁵¹ SLWCD shows a slightly different picture in Figure 17, which indicates that the group that reported both the highest number and the highest rate of these violations in relation to their share of the population was teenaged girls aged 15 to 19. Proportionately, females in the 20 to 34 age groups reported more such violations than younger teens. Few such violations were reported by males. In conclusion, we find that teenaged girls between 15 and 19 were specific targets of sexual slavery. The picture is somewhat different

Figure 17



for reports of sexual abuse, which are presented in Figure 18. The Commission found that, ‘Girls between the ages of 10 and 14 were particularly targeted for sexual abuse.’ This claim is supported by SLWCD: girls 10 to 14 reported more sexual abuse violations, both in terms of raw numbers and proportionately to their share of the population, of all groups. The Commission also interpreted the provision in its mandate, Article 6.2 of the TRC Act, requiring it to pay special attention to the subject of sexual abuses, as referring ‘mainly to women’.⁵² However, as Figure 18 shows, a substantial proportion of males reported sexual abuse, especially the group aged 65 and above. We find that, as with killings, elderly males and females experienced a disproportionate amount of sexual abuse relative to the size of their share of the population. This finding is consistent with the hypothesis that elderly community leaders were the

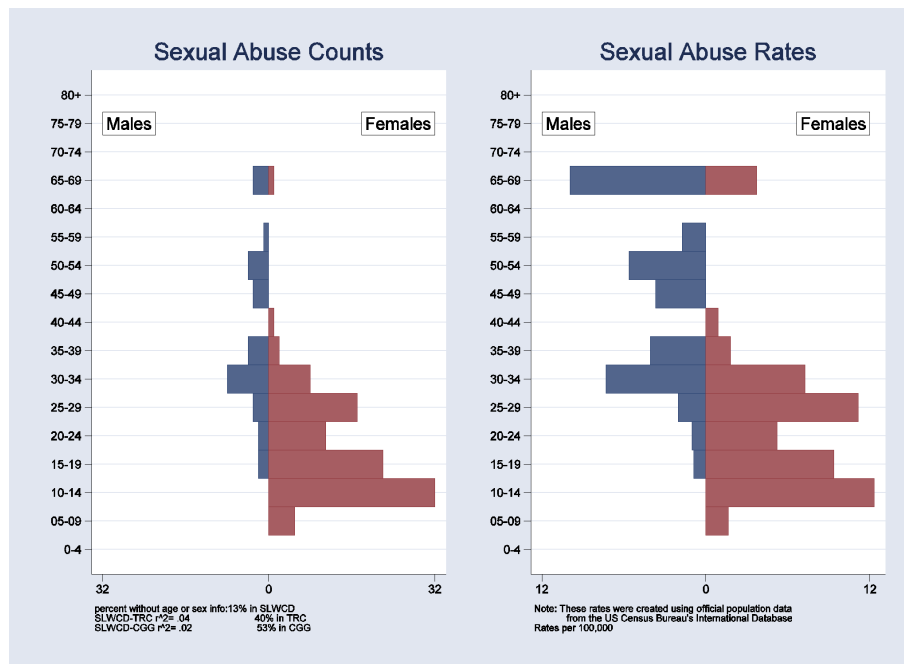
⁴⁹TRC Report, Vol 2, p42, paragraph 133.

⁵⁰HRW, ‘Human Rights Abuses Committed Against Civilians’.

⁵¹TRC Report, Vol 2, p28, paragraph 23.

⁵²TRC Report, Vol 2, p241, paragraph 64.

Figure 18



targets of violations such as killings and sexual abuse, and that perpetrators targeted this subpopulation as a means of intimidating the broader community.

The opposite, however, is seen with respect to rape in Figure 19, which was reported almost only by females, with the highest number of reports, both in raw numbers and proportionately, coming from the 15 to 19 age group. These patterns suggest that acts of rape resulted from substantially different policy and/or practice than killings and sexual abuse. By contrast, forced recruitment was reported predominantly against males. The Commission found that children aged between 10 and 14 were especially targeted for forced recruitment.⁵³ The US Department of State estimated that from 1991 to 1999, the RUF abducted approximately 20,000 persons throughout the country, more than half of whom, primarily children, were released and went through a formal reintegration process.⁵⁴ The UN demobilized 6,845 child combatants between 1999 and 2003⁵⁵, and 6,787 passed through a Disarmament, Demobilization, and Reintegration (DDR) program.⁵⁶ In Figure 20, SLWCD shows a substantially different pattern than that alleged by the groups referenced earlier and that found by the Commission. Proportionately, all age groups between 15 and 39 reported more forced recruitment violations, and in raw numbers, the 15 to 29-year-old age groups all reported more such violations than their younger counterparts. Therefore, we find that young, teenaged boys aged 15 to 19 were approximately equally likely to have been forcibly recruited as males between the ages of 20 and 34. This pattern does not suggest specific policies or practices that explicitly targeted young boys, but rather a general pattern that males aged 15 to 34 suffered approximately equal age-specific risks of forced recruitment, while older and young males, as

⁵³TRC Report, Vol 2, p28, paragraph 23.

⁵⁴US Department of State country report 2003. Web: <http://www.state.gov/g/drl/rls/hrrpt/2003/27750.htm> (2006-02-27).

⁵⁵UN Relief Appeal, 2003.

⁵⁶Jeremy Ginifer, 'Vulnerable and Disadvantaged Groups', in *Armed Violence and Poverty in Sierra Leone*.

Figure 19

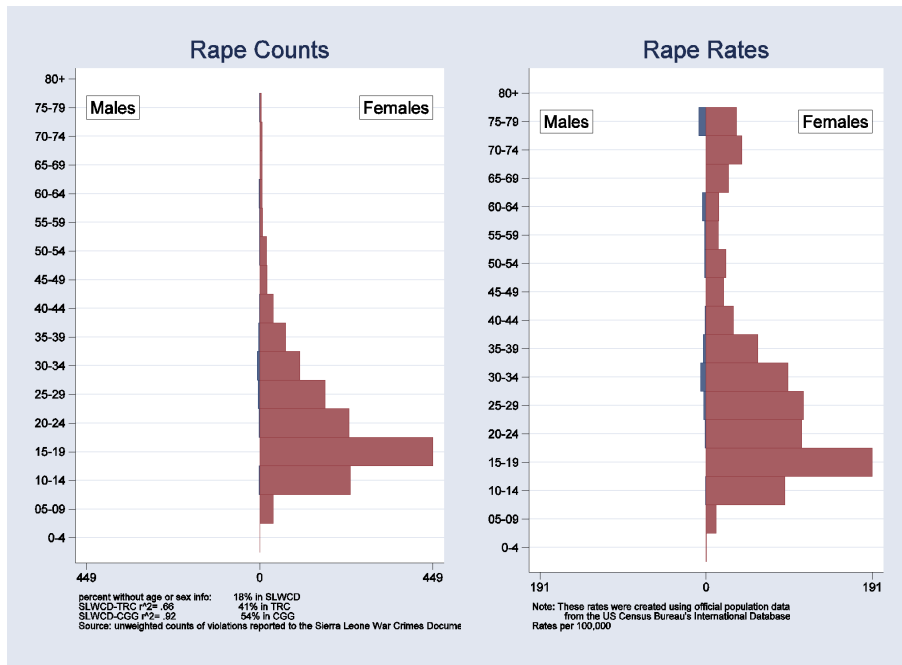
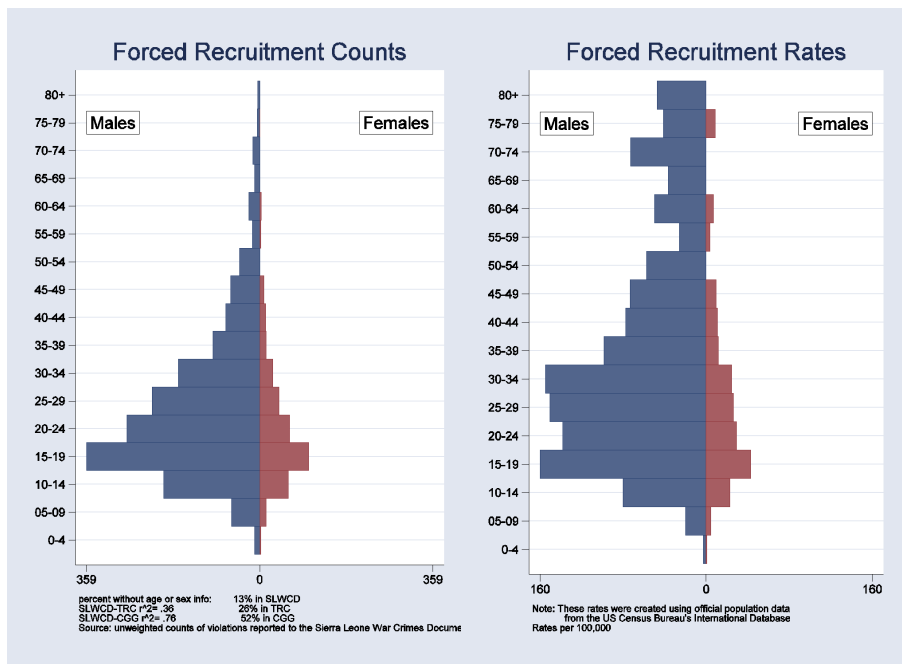


Figure 20



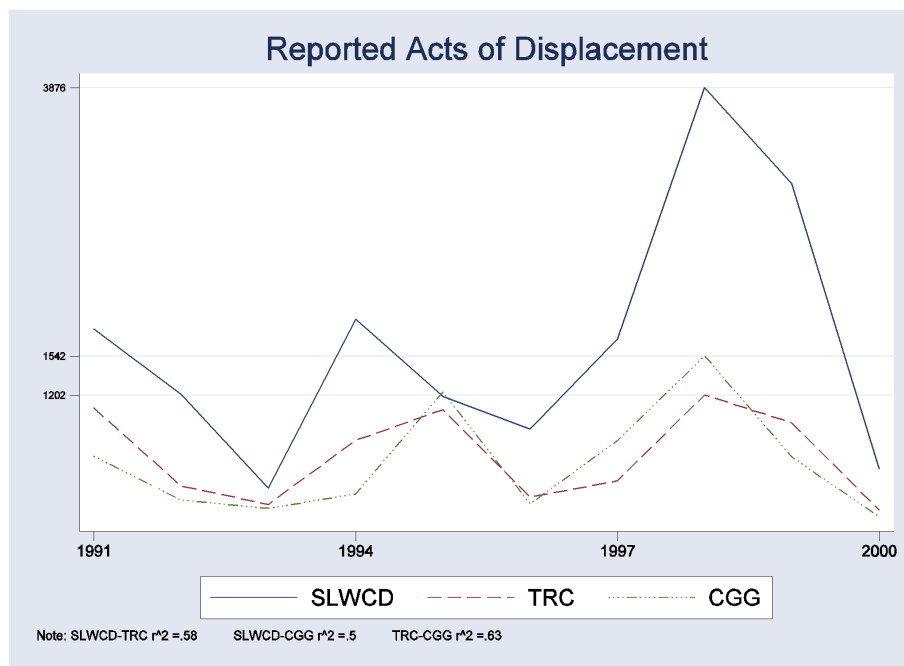
well as females of all ages, had far lower risk.

8 Displacement

The violation suffered by the largest percentage of the population is displacement. Relief organizations estimated that by 1993 about 1 million Sierra Leoneans, out of a total population of 4.5 million, became internally displaced persons (IDPs) within the country or had been forced to take refuge in the neighboring countries of Guinea or Liberia as refugees.⁵⁷ The estimate from the SLWCD is substantially greater: 1.5 million Sierra Leoneans were displaced, as can be seen in . Population movement, overcrowding, and poor sanitation likely contributed to high mortality rates and the spreading of infectious diseases such as malaria, pneumonia, HIV/AIDS, and tuberculosis.⁵⁸ The Commission notes that in many cases displacement, both internal and external, was forced by the destruction of people's homes.⁵⁹

Figure 21 shows the pattern of reported forced displacement for all three datasets across the years of the conflict. The pattern follows the general pattern of phases of the conflict already seen with respect to other types of reported violations.

Figure 21



The number of reported displacements is substantially higher in the third phase than in the first two; the reported displacement in 1991 and 1994 is approximately the same magnitude. The pattern of displacement reported to SLWCD (in the raw counts presented here, and in the estimates presented in) is consistent with the pattern reported to TRC

⁵⁷David Lord, 'The Costs of War', in *The Struggle for Peace and Power in Sierra Leone*. Web: <http://www.c-r.org/accord/s-leone/accord9/intro.shtml> (2006-02-14).

⁵⁸Jeremy Ginifer, 'Vulnerable and Disadvantaged Groups', in *Armed Violence and Poverty in Sierra Leone*.

⁵⁹TRC Report, Vol 3A, p492, paragraph 101.

and CGG, although the magnitudes are different substantially more displacement violations were reported to SLWCD than to TRC or CGG, especially with respect to the third phase of the conflict. As many as 93.7 percent (3375 /

Figure 22

Forced displacement	Destruction of property		
	Not reported	Reported	Total
No displacement	105	121	226
displacement	766	2,609	3,375
Total	871	2,730	3,601

3601) of households interviewed in the SLWCD Survey reported some displacement. Figure 22 shows the number of reported acts of displacement and property destruction. The table shows that destruction of property could have been a risk factor for displacement. While 77 percent (2609 / 3375) of households that reported displacement also reported property destruction, only 53 percent (121 / 226) of the non-displaced households reported property destruction. The households suffering displacement suffered approximately 50 percent more property destruction relative to the non-displaced households. Furthermore, the majority 95 percent (2609/2730) of the households that reported property destruction also reported displacement. By comparison, 88 percent of households that did not report property destruction were displaced.⁶⁰

9 Amputations

Amputations are the single violation that the outside world most often connects with the Sierra Leone conflict. CGG, for example, has highlighted amputations: ‘The special hallmark of both the RUF and the AFRC was the vicious practice of deliberate mutilation, whereby parts of the body inter alia, arms, lips, noses, hands, breasts, and legs were amputated and eyes gouged out. Double arm and leg amputations were also carried out.’⁶¹ Human Rights Watch notes that,

The vast majority of victims are males between the ages of sixteen and forty-five, but women, children, and the elderly are not spared. Men of voting and fighting age are particularly targeted in order to discourage them from giving political or military support to President Kabbah or the Kamajors. When the RUF committed atrocities prior to elections in 1996, they told victims that their hands were being amputated so that they could not vote.⁶²

⁶⁰The relative risk [1.5 = (2609/3375)/(121/226)] that a household was displaced in conjunction with destruction of property is greater than 1, which suggests a positive relationship between displacement and property destruction.

⁶¹Formation of the National Accountability Group, ‘Campaign for Good Governance’. Web: <http://www.slccg.org/nagconcept.htm> (2006-02-15).

⁶²HRW, ‘Human Rights Abuses Committed Against Civilians’, in *Sowing Terror*, July 1998. Web: <http://www.hrw.org/reports98/sierra/Sier988-03.htm> (2006-02-04).

The Commission referred to the practice of amputations as ‘the RUF’s ‘Operation Stop Elections’’.⁶³ In March 2000, the UN’s Humanitarian Coordination Unit reported that the number of survivors of amputation was approximately 600, rather than, as previously estimated, 3,000 to 5,000.⁶⁴ Additionally, in a survey of 240 civilians in Freetown in the first two weeks of May 1999, Médecins San Frontières found that 7 percent had suffered an amputation.⁶⁵

Although reported amputations make up only 1 percent of reported violations, the SLWCD Survey estimates that between 3,000-9,400 amputations were committed during the conflict (see Figure 1). Figure 23 shows the number of reported amputations across the years of the conflict. Where reported killings, displacement, and property destruction showed three broad peaks, reported amputations shows only one, in Phase III of the conflict.

Figure 23

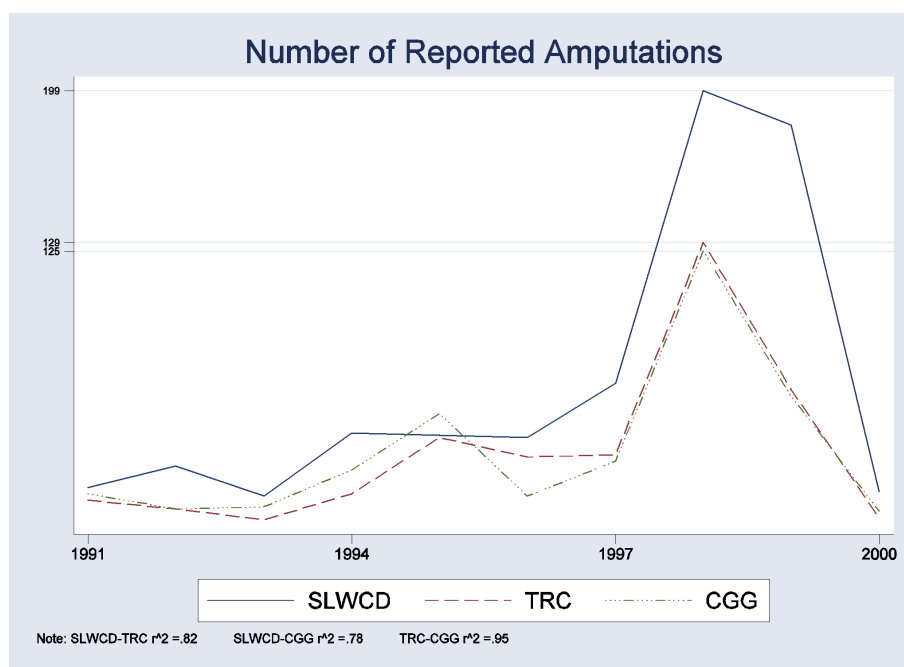


Figure 24 shows that RUF and ‘Rebels’ were responsible for almost all of the small number of reported amputations in the first two phases of the conflict. The AFRC, which is responsible for a significant number of reported amputations in Phase III, did not exist during Phase I, and joined forces with the RUF in Phase III. Figure 25 shows that males were targeted for amputations more than females.

10 Conclusions

The data permit us to make three main findings. First: violence in Sierra Leone was episodic; that is, when violations were large-scale they were concentrated in particular geographical locations, whereas when the scale of violations was

⁶³TRC Report, Vol 2, p44, paragraph 150.

⁶⁴ Cited in David Lord, ‘The Costs of War,’ in *The Struggle for Power and Peace in Sierra Leone*.

⁶⁵ Cited in News Archives, <http://www.sierra-leone.org/slnews0100.html> (2006-02-27).

Figure 24

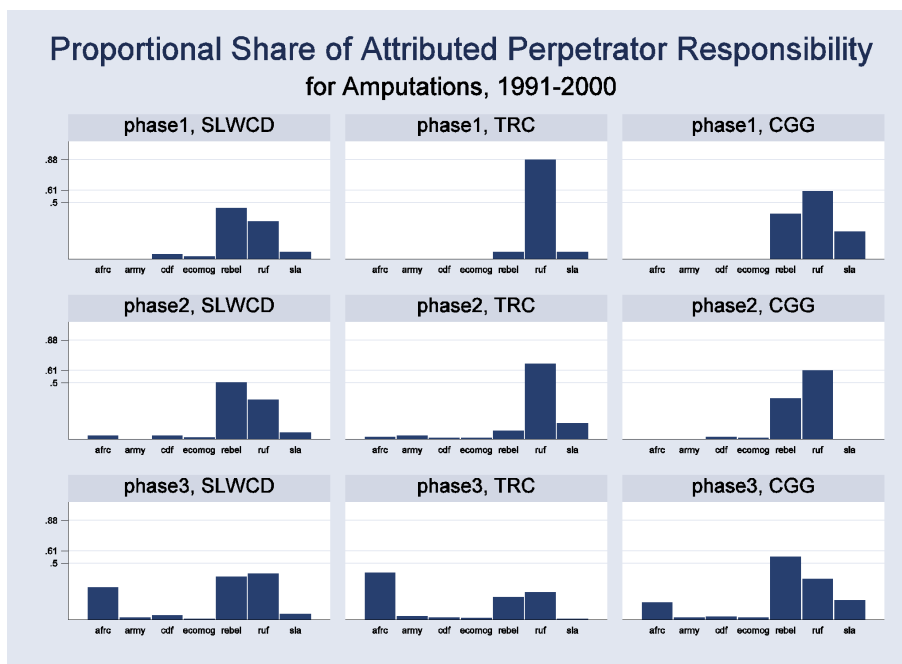
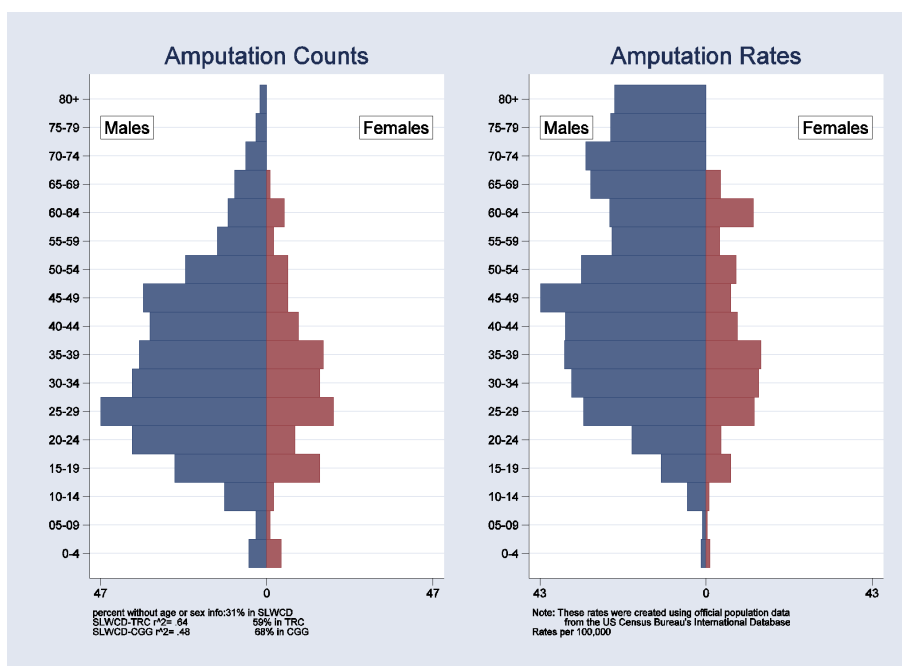


Figure 25



at a low level they tended to be more evenly distributed across space. We believe that the concentrations in specific regions during periods of intensity that is, the episodic nature of the conflict – should open a debate about what the

legal concept of ‘widespread’ means in the context of Sierra Leone because ‘episodic’ and ‘widespread’ would seem to be contradictory ideas.

Second, the households suffering displacement suffered approximately 50 percent more property destruction relative to the non-displaced households, suggesting that property destruction was a risk factor for displacement. Forced displacement and property destruction and theft were by far the most commonly experienced violations during the conflict. An estimated 1.5 million (+/-145,000) people suffered displacement and 430,000 (+/-35,000) people suffered property destruction. We believe that the overwhelmingly most prevalent violations that resulted from the war displacement and property damage have received inadequate analysis from the transitional process. In the context of a country as resource-poor as Sierra Leone, property destruction could quickly lead to elevated mortality, especially for children.

Third, we find that our data are consistent with the Commission’s claim that the RUF was responsible for the majority of killings, property violations, and acts of forced displacement, and further, that the patterns of the RUF’s abuses drove broader patterns in the conflict.

We make the following recommendations for further research. The makeup of ‘rebels’ as a group is still uncertain: what was its command and control structure and its relationship to the RUF? The finding of a relationship between forced displacement and property destruction generates further questions about the practice of these violations, specifically how the different modes of forced displacement were used at different times and places by different perpetrators.

This proportional share of responsibility for human rights violations is not represented in the indictments that have been issued so far by the Special Court for Sierra Leone, which, alongside the TRC, was created in the Lom Peace Accord.⁶⁶ These have been issued against five alleged RUF leaders (two of which have been withdrawn because the leaders in question have died), three alleged CDF leaders, and three AFRC leaders. Two other indictments have been issued, against the former Liberian President Charles Taylor, who is not in the custody of the court, and Johnny Paul Koroma, whose whereabouts are unknown.⁶⁷ With the RUF responsible for as much as three-quarters of the violence of Sierra Leone’s protracted conflict, it seems inconsistent with the mandate to prosecute those ‘who bear the greatest responsibility’ to focus more than half the prosecutions on perpetrating organizations responsible for 5-10 percent of the total violence.

⁶⁶‘The Problem with the Special Court of Sierra Leone,’ by Abdul Karim Bangura, *The Black Commentator*, June 24, 2004. Web: http://www.blackcommentator.com/96/96_sierra_leone.html (2006-02-26).

⁶⁷The Special Court of Sierra Leone. Web: <http://www.sc-sl.org/index.html> (2006-03-07).

11 Bibliography

- Abdullah, Ibrahim, editor. 2004. *Between Democracy and Terror: The Sierra Leone Civil War*. Dakar, Senegal: Council for the Development of Social Science Research in Africa. Amnesty International. 1996. 'Sierra Leone: Toward a Future Founded on Human Rights.' London: Amnesty International. Web: <http://web.amnesty.org/library/Index/ENGAFR510051996?open&of=ENG-2F5> (2006-02-03).
- . 1998. *Sierra Leone: 1998 - a year of atrocities against civilians*. London: Amnesty International. Web: <http://web.amnesty.org/library/index/ENGAFR510221998> (2006-01-26).
- . 2000. 'Sierra Leone: The Human Rights Agenda. Statement by Pierre Sane in Freetown.' Freetown, Sierra Leone: Amnesty International press statement. Web: <http://web.amnesty.org/library/index/ENGAFR510262000> (2006-01-27).
- . 2001. *Guinea and Sierra Leone: No Place of Refuge*. London: Amnesty International. Web: <http://web.amnesty.org/library/Index/ENGAFR050062001?open&of=ENG-SLE> (2006-01-26).
- Ayafor, Martin Chungong, Atabou Bodian, Johan Peleman, Harjit S. Sandhu, and Ian Smillie. 2000. Report of the Panel of Experts Appointed Pursuant to UN Security Council Resolution 1306 (2000), Paragraph 19 in Relation to Sierra Leone. Geneva: United Nations. Web: <http://www.sierra-leone.org/panelreport.html> (2006-02-12).
- Ball, Patrick. 1996. *Who Did What to Whom? Planning and Implementing a Large-Scale Human Rights Data Project*. Washington, DC: American Association for the Advancement of Science. Web: <http://shr.aaas.org/www/contents.html> (2006-01-26).
- , Wendy Betts, Fritz Scheuren, Jana Dudukovich, and Jana Asher. 2002. *Killings and Refugee Flow in Kosovo March June 1999: a Report to the International Criminal Tribunal for the Former Yugoslavia*. Washington, DC: American Association for the Advancement of Science and the American Bar Association Central and East European Law Initiative. Web: http://shr.aaas.org/kosovo/icty_report.pdf (2006-03-05).
- Bausch, Daniel, MD. 1999. 'Sierra Leone.' New York: Doctors for Global Health. Web: <http://www.dghonline.org/n15/hr-sierraleone.html> (2006-02-21).
- Campaign for Good Governance. 2002. 'Formation of the National Accountability Group.' Freetown, Sierra Leone: Campaign for Good Governance. Web: <http://www.slccgg.org/nagconcept.htm> (2006-03-08).
- Central Intelligence Agency. 2006. *The World Factbook*. Washington, DC: Central Intelligence Agency. Web: <http://www.cia.gov/cia/publications/factbook/geos/sl.html> (2006-02-26).
- Commission for Historical Clarification in Guatemala (CEH). 1999. *Guatemala: Memoria del Silencio*. 12 vols. Guatemala: UN Office of Project Services. Web: <http://shr.aaas.org/guatemala/ceh/index.html> (2006-03-05).

'ECOWAS Six-Month Peace Plan for Sierra Leone 23 October 1997 22 April 1998 (Schedule of Implementation).' 1997. Conakry, Guinea. Web: <http://www.sierra-leone.org/conakryaccord.html> (2006-02-24).

Evanson, Elizabeth M. 2004. 'Truth and Justice in Sierra Leone: Coordination between Commission and Court.' *Columbia Law Review*, Vol 104:730, pp730-767. Web: <http://www.columbialawreview.org/pdf/Evenson-WebCopy.pdf> (2006-02-27).

Fanthorpe, Richard. 2003. 'Humanitarian Aid in Post-War Sierra Leone: The Politics of Moral Economy.' London: Humanitarian Policy Group at the Overseas Development Institute. Web: http://www.odi.org.uk/hpg/papers/hpgreport13_b.pdf (2006-02-03).

Gberie, Lansana. 2002. *War and Peace in Sierra Leone: Diamonds, Corruption and the Lebanese Connection*. Ottawa, Ontario, Canada: Partnership Africa Canada. Web: http://www.pacweb.org/e/images/stories/documents/sierraleone2002_e.pdf (2006-01-31).

Ginifer, Jeremy. 2005. *Armed Violence and Poverty in Sierra Leone: A case study for the Armed Violence and Poverty Initiative*. Bradford, West Yorkshire, UK: Centre for International Cooperation and Security Department of Peace Studies. Web: http://www.brad.ac.uk/acad/cics/publications/avpi/AVPI_Sierra_Leone.pdf (2006-01-27).

Human Rights Watch. 1998. *Sowing Terror: Atrocities against Civilians in Sierra Leone*. New York: Human Rights Watch. Web: <http://www.hrw.org/reports98/sierra/Sier988-03.htm> (2006-02-04).

—. 1999. *Forgotten Children of War: Sierra Leonean Refugee Children in Guinea*. New York: Human Rights Watch. Web: <http://www.hrw.org/reports/1999/guinea/> (2006-01-26).

International Crisis Group. 2002. 'Sierra Leone's Truth and Reconciliation Commission: A Fresh Start?' Brussels: International Crisis Group. Web: <http://www.crisisgroup.org/home/index.cfm?id=1801&l=1> (2006-01-26).

—. 2004. *Liberia and Sierra Leone: Rebuilding Failed States*. Brussels: International Crisis Group. Web: <http://www.crisisgroup.org/home/index.cfm?id=3157&l=1> (2006-01-27).

Junger, Sebastian. 2000. 'The Terror of Sierra Leone.' *Vanity Fair*, August, 2000.

Kaplan, Robert D. 1994. 'The Coming Anarchy: How scarcity, crime, overpopulation, tribalism, and disease are rapidly destroying the social fabric of our planet.' *The Atlantic*, February, 1994, Vol. 273, No. 2; pp. 44-76. Web: <http://members.shaw.ca/competitivenessofnations/Anno%20Kaplan.htm> (2006-02-07).

Kenefick, Eric, and Annalisa Conte. 2003. *Sierra Leone: Rural Food Security, Livelihoods and Nutrition Survey and Household Food Security Profiles*. Rome, Italy: Vulnerability Analysis and Mapping, World Food Programme. Web: http://www.daco-sl.org/encyclopedia/4_part/4_2/wfp_sl_vam_report.pdf (2006-02-02).

Lord, David. 2000. 'The Struggle for Power and Peace in Sierra Leone.' Accord 9, September 2000. London: Conciliation Resources. Web: <http://www.c-r.org/s-leone/accord9/intro.shtml> (2006-02-27).

'Peace Agreement between the Government of Sierra Leone and the Revolutionary United Front of Sierra Leone (RUF/SL).' 1996. Abidjan, Ivory Coast. Web: <http://www.sierra-leone.org/abidjanaccord.html> (2006-02-24).

'Peace Agreement between the Government of Sierra Leone and the Revolutionary United Front of Sierra Leone (RUF/SL).' 1999. Lome, Togo. Web: <http://www.sierra-leone.org/lomeaccord.html> (2006-02-24).

Perez, Andrs. 2000. 'Sierra Leone's Diamond Wars: UN Peacekeepers for Rival Gangsters.' Le Monde Diplomatique, June 2000. Web: <http://mondediplo.com/2000/06/02sierraleone> (2006-02-12).

Post-Conflict Reintegration Initiative for Development and Empowerment (PRIDE). 2002. 'Ex-Combatant Views of the Truth and Reconciliation Commission and the Special Court of Sierra Leone in Sierra Leone.' Freetown, Sierra Leone: PRIDE and International Center for Transitional Justice. Web: <http://www.ictj.org/downloads/PRIDE%20report.pdf> (2006-02-02).

Reis, Chen, Lynn L. Amowitz, Kristina Hare Lyons, and Vincent Iacopino. 2002. War-Related Violence in Sierra Leone. Boston and Washington, DC: Physicians for Human Rights and UNAMSIL. Web: http://www.phrusa.org/research/sierra_leone/report.html (2006-02-27).

Resolution 1132 (1997) adopted by the UN Security Council at its 3822nd meeting, October 8, 1997. Web: [http://daccess-ods.un.org/access.nsf/Get?Open&DS=S/RES/1132%20\(1997\)&Lang=E&Area=UNDOC](http://daccess-ods.un.org/access.nsf/Get?Open&DS=S/RES/1132%20(1997)&Lang=E&Area=UNDOC) (2006-02-24).

Revolutionary United Front. 1995. 'Footpaths to Democracy.' The Zogoda, Sierra Leone: Revolutionary United Front. Web: <http://www.sierra-leone.org/footpaths.html> (2006-02-03).

Richards, Paul. 1996. Fighting for the Rain Forest: War, Youth, and Resources in Sierra Leone. London: International African Institute in association with James Curry (Oxford) and Heinemann (Portsmouth, NH).

—. 1999. 'The War and Reconstruction in Sierra Leone.' Development Outreach, Fall 1999. Washington, DC: The World Bank. Web: <http://www1.worldbank.org/devoutreach/fall199/article.asp?id=23> (2006-01-31).

—. 2003. 'The Political Economy of Internal Conflict in Sierra Leone.' The Hague: Netherlands Institute of International Relations "Clingendael" Conflict Research Unit. Web: http://www.clingendael.nl/publications/2003/20030800_cru_working_paper_21.pdf (2006-01-30).

Saloma, Peter. 1999. "Health and human rights in contemporary humanitarian crises: is Kosovo more important than Sierra Leone?" British Medical Journal, 319:1569-1571 (December 11). Web: <http://bmj.bmjournals.com/cgi/content/full/319/7224/1569> (2006-02-02).

Sierra Leone Truth and Reconciliation Commission. 2004. Report of the Sierra Leone Truth and Reconciliation Commission. Accra, Ghana: Graphic Packaging Limited. Web: <http://trcsierraleone.org/drwebsite/publish/index.shtml> (2006-01-26).

Silva, Romesh and Patrick Ball. 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, California: Benetech Initiative. Web: <http://www.hrdag.org/timor> (2006-03-05).

Smillie, Ian, Lansana Gberie, and Ralph Hazelton. 2000. The Heart of the Matter: Sierra Leone, Diamonds and Human Security. Ottawa, Ontario, Canada: Partnership Africa Canada. Web: <http://www.pacweb.org/e/images/stories/documents/heart%20of%20the%20matter.doc> (2006-01-31).

Smillie, Ian, and Lansana Gberie. 2001. 'Dirty Diamonds and Civil Society.' Paper read at CIVICUS meeting, Vancouver, August 2001.

Smith, L. Alison, Catherine Gambette, and Thomas Longley. 2004. Conflict Mapping: Violations of International Humanitarian Law from 1991-2002. Rome, Italy: No Peace Without Justice. Web: http://www.npwj.org/netrep/documents/SierraLeone/NPWJ_CMR_ES_10MAR04.zip (requires registration, 2006-02-09).

United Nations. 2003. 'Interagency Appeal for Relief and Recovery: Sierra Leone.' Geneva: United Nations. Web: <http://www.reliefweb.int/appeals/2003/files/sle03.pdf> (2006-02-12).

United Nations. 2004. 'Focus: AIDS and Conflict: A Growing Problem Worldwide.' In 2004 Report on the Global AIDS Epidemic. Geneva: United Nations Programme on HIV/AIDS (UNAIDS). Web: http://www.unaids.org/bangkok2004/GAR2004_html/GAR2004_12_en.htm#P1704_379265 (2006-01-27).

United Nations Development Program. 2005. Human Development Report 2005. New York: Human Development Reports Office. Web: <http://hdr.undp.org/reports/global/2005/> (2006-30-01).

United States Department of State. 1998. 'Sierra Leone Country Report on Human Rights Practices for 1997.' Washington, DC: United States Department of State. Web: <http://www.sierra-leone.org/usds013098.html> (2006-02-02).

Zack-Williams, Alfred B. 1999. 'Child Soldiers in the Civil War in Sierra Leone.' Bath, UK: Presented at the Development Studies Association Conference. Web: <http://www.devstud.org.uk/publications/papers/conf99/dsaconf99zackwilliams.pdf> (2006-01-27).

Periodicals

Human Development Report (annual). Geneva: United Nations. Web: <http://hdr.undp.org/> (2006-01-27)

A Appendix on Data and Statistical Methods

A.1 Introduction

This appendix describes the data and methods used to produce the analytical findings in this report. It is divided into four main sections

Section A.2 describes the nature, size, and structure of the three datasets that were used in this report. Section A.3 describes the editing, cleaning and name normalization techniques that were applied to the data. Section A.4 presents the various deduplication and record linkage techniques that were used to match multiple reports of the same individual victim. Section A.5 describes the statistical estimation techniques that were used to derive estimates of the total magnitude and pattern of certain violations during the Commission's reference period.

A.2 Data Sources

This report is based on three datasets: TRC, CGG, and SLWCD. TRC and CGG are convenience samples; that is, they were collected from deponents who volunteered to tell their stories. SLWCD is a random sample of households throughout Sierra Leone, and its collection was intended to correct some of the sampling limitations associated with convenience samples, namely that the analysis derived from them cannot reliably be projected across the rest of the population.

A.2.1 Narrative Testimony Data Sources

The data model for TRC and CGG is that described in *Who Did What to Whom?*⁶⁸ When gathering testimony, it is important to preserve the complexity of the data. In database terms, each story is structurally complex because of the number of variables involved. Each narrative may contain from one to many victims, violations, and perpetrators, and all of these elements may be interrelated in multiple directions. If this structural complexity is not respected in the database representation, the resulting statistics will not be representative of the stories from which the quantitative data were derived

The basic elements of a human rights narrative are: many victims, many violations, and many perpetrators. A deponent may speak about violations that happened to one victim, or that happened to many victims. A deponent's story may discuss only her own detention and subsequent torture; or she may also speak about her son's killing and her husband's disappearance; or she herself may or may not be a victim. In addition, each of the victims described in the statement may have suffered one or more gross violations. The witness's son may have been detained and tortured on several separate occasions before he was killed. These violations may have been connected to others that occurred at the same time and in the same place. Or they may have been isolated incidents.

Each violation may have been committed by one or more perpetrators, who may or may not be identifiable (and who themselves may have been victims in the past). The witness may or may not have seen the violation occur and may or may not be able to identify the perpetrators. For example, she may have been notified that her son's body was

⁶⁸Ball, Patrick: *Who Did What to Whom? Planning and Implementing a Large-Scale Human Rights Data Project*. (1996), AAAS: Washington, DC, USA.

found but know nothing more. If the witness was herself a victim, she may be able to describe the organization to which the perpetrators of her violations belonged or have recognized its identity. She may also have personally recognized one or more of the perpetrators. Furthermore, each of the perpetrators identified in the narrative may have been responsible for one or more violations. For example, the witness may identify a single individual who was responsible for both her own torture and her son's killing.

A.3 Truth & Reconciliation Commission Dataset

A.3.1 Data Collection

The TRC database consists of 7,706 depositions cataloging 40,242 violations collected between 2002 and March 2003. Statements were taken from voluntary deponents in 141 of the 149 chiefdoms as well as in Gambia, Guinea, and Nigeria where refugees from Sierra Leone were living.⁶⁹ The statements they gave offer detailed insight into the experience of particular victims and perpetrators, and every statement therefore deserves careful study.⁷⁰

A.3.2 Data Coding

Data coding is the process of transforming unstructured narrative information on violations, victims, and perpetrators into a countable set of data elements, without discarding important information or misrepresenting the collected information. The TRC narrative statements were coded into countable units according to the aforementioned *Who Did What to Whom?* data model. Acts were classified into the following fourteen violation types using a controlled vocabulary designed to ensure the consistent application of these definitions⁷¹:

- Abduction
- Amputation
- Arbitrary Detention
- Assault/Beating
- Destruction of Property
- Drugging
- Extortion

⁶⁹For a detailed description of this dataset, refer to Conibere, R. et al. 'Statistical Appendix to the Report of the Truth and Reconciliation Commission of Sierra Leone: A Report by the Benetech Human Rights Data Analysis Group and the American Bar Association Central European and Eurasian Law Initiative to the Truth and Reconciliation Commission. (5 October, 2004)

⁷⁰Truth and Reconciliation Commission of Sierra Leone & Benetech Human Rights Data Analysis Group. 'Sierra Leone- Truth and Reconciliation Commission Data.' published 21 December, 2005. The anonymized data are freely available at www.hrdag.org/resources/SL-TRC_data.html (2006-03-07).

⁷¹For details on the design and implementation of controlled vocabularies for human rights data projects, see Silva, Romesh 'On the Maintenance and Measurement of Inter-Rater Reliability when Documenting Large-Scale Human Rights Violations.' *Proceedings of the Joint Statistical Meetings of the American Statistical Association, the International Biometric Society (ENAR and WNAR), the Institute of Mathematical Statistics, and the Statistical Society of Canada.* August, 2002.

- Forced Cannibalism
- Forced Displacement
- Forced labor
- Forced Recruitment
- Killing
- Looting
- Physical Torture
- Rape
- Sexual Abuse
- Sexual Slavery

Actor role affiliations for ‘perpetrator’ and ‘victim’ records were classified into eleven specific institutions: RUF, AFRC, SLF, CDF, ECOMOG, GAF, ULIMO, Police, AFRC/SLA, Miscellaneous, Rebels.

A.3.3 Data Collection and Data Coding Results

Of the 9,063 statements collected by the TRC, 36.9 percent (3347/9063) were collected in the Northern Region, 25.2 percent (2280/9063) were collected in the Southern Region, 19.9 percent (1802/9063) in the Eastern Region and 15.0 percent (1357/9063) in the Western Region.

As is shown in Table A1, the most frequently reported violations during the TRC’s statement-taking process included forced displacement 19.8 percent(7983/40242), abduction 14.8 percent (5968/40242), arbitrary detention 12.0 percent (4835/40242) and killing 11.2 percent (4514/40242).

A.4 Campaign for Good Governance Dataset

A.4.1 Data Collection

The CGG data collection project collected 2,788 statements that documented 25,447 violations against 14,051 victims.

The CGG dataset began in 2002 as a preliminary investigation known as the ‘Mapping Project’ carried out by the CGG under contract to the Office of the United Nations High Commissioner on Human Rights (OHCHR). At that stage, the project took 1,316 statements at locations throughout the country. Many of the preliminary conclusions reached by the Mapping Project were later borne out by the TRC. The CGG followed up by collecting an additional 1,472 interviews between January and September 2004.

Figure 26: Count of Violations by Violation Type, 1991-2000

Violation Type	Percent of violations	Count of violations
Forced Displacement	19.8	7983
Abduction	14.8	5968
Arbitrary Detention	12	4835
Killing	11.2	4514
Destruction of Property	8.5	3404
Assault / Beating	8.1	3246
Looting of Goods	7.6	3044
Physical Torture	5.1	2051
Forced Labor	4.6	1834
Extortion	3.2	1273
Rape	1.6	626
Sexual Abuse	1.2	486
Amputation	0.9	378
Forced Recruitment	0.8	331
Sexual Slavery	0.5	191
Drugging	0.1	59
Forced Cannibalism	0	19
Total		40242

Source: Sierra Leone Truth and Reconciliation Commission's Database

A.4.2 Data Coding

The CGG dataset was coded according to the same coding frame as the TRC narrative testimony data (see above for details).

A.4.3 Data Collection and Data Coding Results

In the 2,788 statements collected by the CGG, of the 25,447 violations documented, 25.2 percent (6407/25447) forced displacements were reported, 15.2 percent (3869/25447) killings, 14.1 percent (3584/25447) abductions and 8.2 percent (2090/25447) acts of destruction of property. This is shown in Table 27, below.

Figure 27: Count of Violations by Violation Type, 1991-2000

Violation	Count	Percentage
Forced Displacement	6407	25.2
Killing	3869	15.2
Abduction	3584	14.1
Destruction of Property	2090	8.2
Looting of Goods	1661	6.5
Assault/Beating	1656	6.5
Arbitrary Detention	1419	5.6
Forced Labor	1206	4.7
Physical Torture	1047	4.1
Rape	711	2.8
Extortion	651	2.6
Amputation	343	1.3
Sexual Abuse	297	1.2
Forced Recruitment	247	1.0
Sexual Slavery	204	0.8
Drugging	46	0.2
Forced Cannibalism	8	0.0
Other	2	100
Total	25447	

Source: Sierra Leone Campaign for Good Governance's Database

A.5 Sierra Leone War Crimes Documentation Survey (SLWCD)

The Sierra Leone War Crimes Documentation Survey (SLWCD) is made up of structured interviews with the heads of households drawn from a probability-based random sample of households throughout Sierra Leone's thirteen districts.

A.5.1 Sample Design

To draw the sample for SLWCD, the project used area sampling, a method that is often used when a complete register of the population of interest is not available. Area sampling involves dividing the total area under investigation into small sub-areas that are then sampled either randomly or through some restricted random process. Each of the chosen sub-areas is then fully inspected and enumerated and may in turn become a frame for further sampling, if desired.⁷²

The sample design for SLWCD was based on 2,522 maps developed by the Sierra Leone Central Statistics Office (CSO) during the 1985 Population and Housing Census in Sierra Leone, the most recent such data available. Each map delineated an enumeration area of approximately 100 households. The maps were stratified by rural and urban regions.⁷³ The ‘rural’ areas were further stratified by chiefdom, and the ‘urban’ areas were further stratified by the current population size of each enumeration area according to official CSO projections.

Budgetary considerations limited the project to 600 enumeration areas. These were sampled in proportion to the size of their population: 407 enumeration areas were rural and 193 were urban. Because the rural areas were further stratified by chiefdom, the project planned to sample a proportional number of rural enumeration areas, via simple random sampling, based on chiefdom population size. The 193 urban enumeration areas were sampled via Simple Random Sampling within strata; the number of urban enumeration areas sampled within each stratum was based on the population size of that stratum.

Within each sampled rural enumeration area, the project consulted local experts to check that the map was still accurate, and revised the maps if necessary to include any new villages that had formed as well as remove any that no longer existed. The project then used a random sampling technique to pick one village within the enumeration area. The number of households in each sampled village was determined in consultation with the village chief. From the list of households, six were randomly chosen for interviews. The procedure for selecting households in urban enumeration areas was quite different. The maps for these areas contained diagrams of the housing units and streets. After revising the elements of the maps that had changed since 1985, interviewers would follow a ‘random walk’ through the enumeration area until they had obtained responses from six households.

A.5.2 Questionnaire design

The SLWCD questionnaire contained nine registers:

1. a household register that enumerated all members of the household, and
2. eight violation registers based on broad categories of abuse phenomena.⁷⁴

With each violation register, the interviewer began by asking a question about that specific category of abuse, and recorded all cases of that abuse known to the respondent regardless of whether the victim of the abuse being reported

⁷²For a definition of area sampling, see International Statistical Institute, *The Oxford Dictionary of Statistical Terms*, edited by Yadolah Dodge, Oxford University Press, 2003.

⁷³The definition of ‘urban’ and ‘rural’ was based on the Sierra Leonean National Statistical Office’s definition. This definition had been constructed based on the observed population density in 1985.

⁷⁴The eight abuse categories were physical abuse, property damage, forced displacement, drugging, abduction and imprisonment, forced labor, sexual violence and forced cannibalism. For the purposes of the survey, killing was not conceived of as a human rights abuse in its own right. Instead, it was characterized as a (potential) result of a given abuse phenomenon.

was a resident of the sampled household. Only one deponent was interviewed for each sampled dwelling.

The questionnaire was developed in English. It was then translated into Sierra Leone's six most frequently spoken languages (Krio, Kono, Koranko, Temne, Limba and Mende). It was then translated back into English in order to ensure that the different versions of the survey were all consistent. No substantive differences were detected during the back-translation exercise; however, a later language review step was required.

The SLWCD questionnaire was then tested in all six languages using standard cognitive interviewing techniques.⁷⁵ The cognitive interviewing was performed in four waves; after each wave the questionnaire was modified in all of the languages and retested. After the cognitive interviewing process, during a language review exercise, minor modifications to the English version of the questionnaire became necessary in order to ensure that all six local language versions of the questionnaire were conceptually equivalent.

A.5.3 Survey Implementation

The project conducted a public education campaign in rural areas to help develop public awareness of the survey's objectives and strategy, and also to elicit support from local officials and village elders. Small advance teams consulted with local officials and village elders in enumeration areas in advance of the enumeration teams. In Freetown, local television and local radio shows spots informed the public about the survey.

All forms of public outreach stressed that the project was independent of the proceedings of the Special Court of Sierra Leone and also that the household/deponent selection process was to be based on probability-based random sampling.

A.5.4 Data Collection Results

The SLWCD data collection was carried out over a nine-month period starting in January 2004. The enumeration team surveyed 3,633 heads of household in 3,631 dwellings in all 150 chiefdoms of Sierra Leone as well as the Western Region. The SLWCD database contains 64,717 total records of violations, 11,909 of which are resident violations and 47,629 are non-resident.

A.5.5 Methodological Description of Data Editing, Cleaning & Name Normalization Techniques

A considerable amount of data cleaning, editing and preparation was carried out on the TRC, CGG, and SLWCD datasets to get them ready for the record linkage process. This section provides a brief overview of the various procedures we used.

⁷⁵See Willis, Gordon B. 1999. Cognitive Interviewing: A 'How To' Guide. Web:<http://appliedresearch.cancer.gov/areas/cognitive/interview.pdf> (2006-03-07).

A.5.6 Name Cleaning & Canonicalization

We cleaned the firstname and lastname information for each identified victim of punctuation marks, leading and ending spaces and shifted names into their correct fields. Then each name component (i.e. Firstname and lastname) was canonicalized.⁷⁶ Person names contained a significant amount of variation in the spellings and in punctuation. Name variation has many causes. In open-ended narrative statements, such as the TRC and CGG data collection processes, the deponent may be a close relative, friend, neighbor or distant acquaintance of the victim, and he or she may or may not know how to spell the names of the reported victim. Transcription by the statement-taker may involve application of additional spelling and punctuation rules and even incorporate spelling errors. Similarly, spelling and punctuation transformations may take place at the data coding and data-entry stages. For example, the following lastnames were canonicalized to the unique lastname ‘Abdulai’:

- Abdulah
- Abdulai
- Abdulia
- Abdulie
- Abdullah
- Abdullai

A.5.7 Duration Code Cleaning

Duration information, documenting the length of time for particular violations, were collected using the SLWCD questionnaire. The duration codes used for this information were of alphanumeric format, with the non-numeric part indicating whether the duration unit was hours, days, weeks or months.

The project developed parsing code, which cleaned these duration values, by replacing invalid codes with null values and by mapping all violation codes into a standardized unit of time for this analysis. The standardized unit of time used was ‘number of days’.

A.5.8 Violation Code Cleaning

All three data collection projects focused their data collection around the documenting of human rights violations. Of the 40,242 violation records documented in the TRC database and the 25,447 violation records documented in the CGG database, only one invalid code was documented. Of the 64,718 violation records documented in the SLWCD database, 1.1 percent (684/64,718)

⁷⁶Canonicalization is a process of reducing each name to the simplest and most significant form possible, without loss of generality.

did not contain at least one raw, valid violation code. 86.3 percent (590/684) of the invalid violation codes were able to be recoded into valid codes for this analysis.

A.6 Standardization of Victim Death Codes

As mentioned in Section A.5.2, killing was not conceived as a human rights abuse in its own right during the SLWCD Survey. Instead, it was characterized as a (potential) result of a given abuse phenomenon. 5.2 percent (3384/64718) of raw victim death codes were invalid, of which 42.8 percent (1450/3384) were able to be recoded into valid codes for this analysis.

A.7 Record Linkage Overview

Record linkage is the process of identifying, linking and merging records from one or many datasets which represents the same person (or individual victim, in the context of human rights data projects). This process is important for two particular reasons: (i) in order to ensure that any ensuing descriptive statistical analysis does not double-count the same violation and (ii) to identify the overlap patterns in reporting across the different datasets as a necessary input for multiple systems estimation.

There are two categories of record linkage: de-duplication (i.e. identification of duplicate records within a given datasets), and linkage between datasets (i.e. Identification of duplicate records across different datasets). In this project, though, we combined these two processes into one process. This section presents the individual steps involved in identifying records about the same victim within and across the three data projects: in particular, how the feasible set for linkable records is defined, the process by which the link pairs are generated and weighted, the clustering of records into groups of duplicates, and the merging of duplicate records into representative records.

A.7.1 Defining the Feasible Set of Records for Record Linkage

Actor rejection rules were defined to identify the feasible set of records on which MSE estimates could be made. The following actor rejection rules were defined:

- violation type is either ‘Killing’ or ‘Amputation’
- drop representative records with no dated violation (year-only violation dates are feasible)
- drop representative records where both the firstname and lastname are less than or equal to one character long

Actors who failed any of these tests were excluded from the linkage process and thus from subsequent analysis.

A.7.2 Defining Link Pair Rejection Rules

After the feasible set of records for record linkage were defined, we developed criteria to identify pairs of records which it would be meaningful to link within or between datasets (i.e., records for which there was a basis to claim that they represented the same person).

The following link-pair candidate rules, which identify actors which should not be linked, were applied to the data:

- the actors were mentioned in the same survey, the same incident or the same act (i.e. They have the same folder_id, case_id or act_id (and both are non-null value)),
- both victim records were residents in the SLWCD database,
- a fatal SLWCD-non-resident record can be linked to a SLWCD-resident record,
- for birth dates: both dates are ‘perfect dates’, the number of years differ by more than 5 years, and day and month values do not match
- for fatal violation dates: both dates are ‘perfect dates’, the number of years differ by more than 3 years, and day and month values do not match
- sex mismatch (if both sex values are non-null)

To look for duplicates, we consider every possible pair of victims taken two at a time (where order doesn’t matter). We record each pair of victims that has identical, non-blank values in any two of the following columns:

- firstname
- lastname
- birth_year
- death_geo1
- death_geo2
- death_geo3
- death_year
- death_month
- death_day

Each pair of records that compares equal on at least two columns is used to construct a link pair, which contains the actor_id of each record plus a weight that is approximately $1/\text{probability-of-coincidence}$, where ‘probability of coincidence’ is the probability that two records have the same values in some columns ‘by coincidence.’ The probability-of-coincidence denominator is the record count, and the numerator is the frequency of the in-common field values found in a collection of columns (called a ‘chunk’):

- name: firstname, lastname
- birth_date: birth_year, birth_month, birth_day

- death_date: death_year, death_month, death_day
- death_loc : death_geo1, death_geo2, death_geo3

If two fatal victims only match on death_year+death_month but not death_day, the numerator is the frequency of their in-common values for these two columns (which is ~ 30 times higher than the frequency of either victim's death_year+death_month+death_day, since there are ~ 30 days/month). If two victims have (possibly-partial) matches on name and death_loc for example, we multiply the name weight times the death_loc weight to get an aggregate weight; this aggregate weight is 'over-optimistic', because the name-weight and the death_loc-weight may be statistically dependent (because name frequencies will depend on ethnicity, and ethnic demographics will depend on geography).

We take the weighted link-pairs from the previous step, and cluster them into a multiplicity of 'link group sets' (sets of groups of 'duplicates'), where each set has a specific minimum weight that is required between all possible pairs of victim actor_id values that are allowed to be in the same 'link group' (i.e. group of actor_id values considered duplicates of each other). Each 'link group' is given an arbitrary (but unique) group_id.

The 'link group sets' from the previous step are applied one-by-one to the merged database of victim-role records. For each set of victim-roles with the same group_id, a 'representative record' is constructed. When merging a set of duplicate records to form the 'representative record', for each column, we choose the modal (most-frequent) unique value amongst all the records in that group. Note that the blank-value unique (that is, empty field) is considered under-reporting, and is ignored. If there are N different uniques with the same frequency, an N-sided coin is flipped to select one of them.

A.8 Statistical Estimation Techniques used in the Analysis of Fatal Violations and Displacements

Both survey estimation and multiple systems estimation (MSE) methods were used to make estimates of the total magnitude and pattern of violations between 1991 and 2000. This section describes the methodological basis for both of these methods. We present, in detail, how the survey weights were calculated for the SLWCD-based survey projections. We also present the mathematical basis for the MSE estimates.

A.8.1 SLWCD Survey Weight Calculations

The variables already in the database were not sufficient to prepare the weights. Several pieces of additional information needed to be added to the database, and multiple new variables were formed. In order to explain the process as clearly as possible, we will first list all of the variables, what they represent, and how they were formed. We will then review the process by which the two final weights the household-level and individual-level weights (HW and IW) were formed.

A.8.2 Variables Created

1. EAW (enumeration area weight) This weight is equivalent to the number of enumeration areas in the stratum that a single sampled enumeration area represents. In other words, it is the number of EAs in the stratum in question, divided by the number of EAs in the stratum's sampling frame. This weight is not intended to be used for direct estimation purposes, and is calculated at the stratum level, meaning it is identical for all records from the same stratum.
2. EARF (enumeration area raking factor) Because the urban enumeration areas were stratified within region and population size categories, not within district, the sum of the EAWs for the district is not equivalent to the number of urban EAs in the district. This is corrected by EARF, which is a raking factor calculated by taking the sum of the EA weights for the district and dividing by the number of urban enumeration areas for that district. This raking procedure ensures two things: one, that the relative probabilities of selection for the EAs to remain proportional to each other; and two, that the sum of the EA weights is correct. This variable is set to one for the rural enumeration areas.
3. UEAW (final urban enumeration area weights) This is EAW multiplied by EARF.
4. VWEAW (Village within EA weights) Used only for the rural enumeration areas, and identical for all records from the same enumeration area. For the urban EAs, this weight is set to one. For most rural EAs it is set to the number of villages in the enumeration area. There are some exceptions, listed below:
 - In the Dibia and T.M.S. Chiefdoms of Port Loko (2409 and 2410), the estimated population size was very high. As a result, in those chiefdoms the number of enumeration areas to be randomly chosen was greater than the number of enumeration areas in the chiefdom. To resolve this issue, enumeration areas were picked 'with replacement', and the number of times a given enumeration area was randomly selected to be in the sample determined the number of villages selected from that area. In these cases, the EAW for the enumeration area is set to 1 (it is self-representing), and VWEAW is set to the number of villages in the enumeration area divided by the number of villages selected in the sample.
 - In the Kaffu Bullom Chiefdom of Port Loko (2403), two rural enumeration areas were selected out of sixteen. When the interviewers arrived at the chiefdom, the chief insisted that more villages be interviewed. He provided a list of all of the villages in his chiefdom. The two villages that had already been sampled were removed from the list, and out of the remaining 74 villages three more were randomly selected. Individual village weights (VWOs) for these five villages are calculated as follows. First, the VWOs for each of the two sets of villages is calculated as if the other set did not exist:

$$\text{EA-based selection: } \frac{16}{2} \times 7 = 56$$

$$\text{List-based selection: } \frac{74}{3} = 24\frac{2}{3}$$

Then a 'raking factor' is developed to allow the weights to add up to 76, the total number of villages on the chiefdom list:

$$\text{Sum of weights: } 56 + 56 + 24\frac{2}{3} + 24\frac{2}{3} + 24\frac{2}{3} = 186$$

$$\text{Raking factor: } \frac{76}{186}$$

Finally, the raking factor is multiplied by the individual weights.

- In 13 cases, the original map used in the field was lost. In these cases, the VWEAW is replaced by the average size of the other villages sampled within the district. Ten of the missing maps were from Bo, which has a rural EA sample size of 16. The other three cases were in Kenema (one of 36 rural EA maps missing), Port Loko (one of 108 rural EA maps missing) and Moyamba (one of 28 rural EA maps missing).
 - In the Gbanti Kamaranka and Magbaiamba chiefdoms of the Bombali district (2103 and 2105), the interviewers discovered only after they had arrived in the field that they had incorrect or deficient maps, so they selected one village at random from the entire chiefdom. In this case, the 'EA' probability is 1, and the overall HH weight is formed directly using the 2004 Census HH count for the chiefdom. No VWEAW is formed.
 - Finally, in seven cases the interviewers in the field selected more than one village to be part of the sample, for a variety of reasons. There were five such enumeration areas in the Kono district, one in the Kenema district, and one in the rural Western area. In these cases, VWEAW is set to the number of villages in the enumeration area divided by the number of villages from the EA selected in the sample.
5. VWO (Village weight old) UEAW multiplied by VWEAW. This represents a 'village' weight for the rural EAS, and is just UEAW for the urban EAS. Please note that attempting to create estimates using 'village-level' weights is not recommended.
 6. HUC (Housing unit count) A count of the housing units (HUs) sampled within an urban EA/village. See the MHAF explanation for why this count is needed. In most cases, this is equivalent to the count of the households for the urban EA/village.
 7. MHAF (Multiple household adjustment factor) In urban areas, a map was used to determine which houses to visit. In some of the rural enumeration areas it was not possible to contact the village chief and gain a list of all the households within the village or town. For example, many of the 'rural' villages contained 50 households or more. In these cases, the interviewers recreated the selection method for the urban enumeration areas by developing their own map of the region, numbering the housing units, and then randomly selecting dwellings to visit. The difficulty in both of these cases – urban enumeration areas and large rural villages – is that a single dwelling might house multiple households. To compensate for this, the fieldworkers were instructed to interview each household separately and mark their interview forms with 'A' and 'B' to distinguish the households. Because of this, we have a record of the number of multiple-household dwellings encountered during the survey. The multiple household adjustment factor makes use of this information. Let CR be the count of interviews conducted in the urban EAS and in those rural EAS that have a final household count (FHC, explained below) of 50 or more within a single geographic region R (east, west, north, or south). Then the MHAF is the count of housing units represented by the interview multiplied by the ratio of CR to the count of interviews in the same EAS represented in CR. In other words, it is the percentage of the household weight that represents a count of housing units, which were the actual sampling units in the urban enumeration areas. MHAF is set to one for villages in rural EAS where the household count for the village is less than 50.
 8. PHC (Preliminary household/housing unit count) This variable is complicated, and is identical across records from the same urban EA, or from the same village in a rural EA. For urban EAS and rural EAS where large villages were picked (50 housing units or greater), PHC is set to the number of housing units on the EA map,

if that map is available. For rural EAS where small villages were picked (under 50 housing units), PHC is set to the number of households in the village selected, if that number is available. Unfortunately, twelve of the original maps, as well as many of the household count records, were lost. In these cases, PHC is set to the maximum housing unit identification number (see explanation below). Then the variable FHC (final household/housing unit count) is set to the actual count of households/housing units if it is available, or an estimated household/housing unit count (HHC), based on the value of PHC, if it is not. The FHC estimation procedure is detailed below. Please note that two EAS, 0304-04 and 0305-05, were put together as one unit map; both EAS were in the sample, so we obtained 12 interviews from the areas combined. Therefore PHC for these EAS is the sum of their individual housing unit counts divided by 2.

9. HHC (Household/housing unit count) The estimated household/housing unit count. See FHC for an explanation.
10. FHC (Final household/housing unit count) The actual household/housing unit count when it is available, otherwise the estimated household/housing unit count.

Estimation Procedure Please note that households/housing units within a village/urban EA were drawn via a simple random sample; each household/housing unit was assigned a number between 1 and a , where a is the total number of households/housing units in the village/enumeration area. Next, n households/housing units were randomly picked based on their assigned number. The unique identification code for each interview was created using the number assigned to the household/housing unit during the random selection. As such, we can consider the household/housing unit ‘code’ a random variable from a discrete uniform distribution supported by the integers between 1 and a .

Now, let:

$$y = \max(X_1, \dots, X_n), n \leq y \leq a$$

then it follows that: $g(y) = \frac{\binom{y-1}{n-1}}{\binom{a-1}{n-1}} n \leq y \leq a$

The expected value for y is then: $\frac{(a-1)n}{n+1}$

Therefore, an unbiased estimator of a is given by: $\frac{n+1}{n}y - 1$

In cases where it is not available, the count of household/housing units is estimated by substituting for y the maximum of the assigned household/housing unit numbers for the village/urban EA, and substituting for n the number of households/housing units in which surveys were conducted in the village/urban EA. Where there is additional variability from the survey data because the estimates have been formed by approximating the household/housing unit counts, we accounted for it by using jackknifing to estimate the standard errors.

11. VCAI (Village count approximated indicator) Set to ‘*’ if the village count for the rural EA was approximated as above.
12. HCAI (Household count approximated indicator) Set to ‘*’ if the household count for the village or urban EA was approximated as above.
13. HHVO (Household/Housing unit within village/urban EA weight old) - FHC divided by the number of housing

units/households for the urban EA/village. In other words, the number of housing units/households each sample unit represents.

14. HHWO (Household/Housing unit weight old) The raw household-level weight, before the multiple household adjustment factor, non-household buildings adjustment, and raking to the 2004 census results.
15. CONHB (Count of non-household buildings) As they reached households in the prearranged random order, interviewers in the urban enumeration areas also found destroyed and burned-out buildings. If these were destroyed or burned-out households, the interviewers were instructed to note them as such on the map and continue to the next sample unit. Similarly, if they found a commercial building instead of a house, they were to mark it as such on the map and continue. This variable is a count of the number of 'non-household' buildings encountered by the interviewer in the enumeration area. It includes both burned-out or destroyed houses and commercial units.
16. NHBA (Non-household buildings adjustment) The CONHB variable can be used to estimate the percentage of buildings within an urban EA/large village that are not housing units. Let q EF, CF, WF, Eastern, Southern, Northern represent one of the three areas of Freetown or one of the three non-western regions of Sierra Leone. Let iq indicate an urban enumeration area, or village with FHC of 50 or greater, within the area q , and qi indicate an area that contains enumeration area/village i . To create the non-household buildings adjustment (NHBA), we note the percentage of such buildings encountered by the interviewers out of all buildings canvassed, including non response households. EAs for which NHB counts have been lost are not included in these calculations. Estimates of this percentage based on individual enumeration areas would be highly variable, hence we borrow strength across multiple enumeration areas, working on the assumption that the EAs will have similar composition within each of the six areas of q . The formula for NHBA is as follows (HU = housing unit):

$$NHBA_{qi} = \frac{\sum \#of\ respondent\ HUs\ in\ EA_{iq} + \sum \#of\ non-response\ HUs\ in\ EA_{iq}}{\sum \#of\ respondent\ HUs\ in\ EA_{iq} + \sum \#of\ non-response\ HUs\ in\ EA_{iq} + \sum \#NHBs\ in\ EA_{iq}}$$

17. PUHW (Provisional urban household weights) The adjusted Household-level weight, before raking to the 2004 census. HHOW multiplied by MHAF and NHBA.
18. HHRF (Household raking factor) the raking factor to bring the household weights in line with the 2004 census counts of households for chiefdoms. It is simply the count of households for the chiefdom from the 2004 census divided by the sum of PUHW for the chiefdom.
19. HW (Household weight) The final household weight; this is the variable to be used in estimation for households.
20. IWHC (Individual within household count) Set to the number of individuals listed on the household roster; this is a household-level variable.
21. IRF (Individual raking factor) - The raking factor to bring individual weights in line with the 2004 census.
22. IW (Individual weight) The final individual weight; this is the variable to be used in estimation for individuals.
23. COR (Count of refusals) During the fieldwork in the urban enumeration areas, the interviewers were instructed to follow a prearranged order in attempting to conduct interviews at households until they had successfully

conducted three interviews. If a household refused, they were to mark it on their map. After the fieldwork was completed, this variable was created from the number of refusals listed on the maps.

24. COON (Count of other non-response) In some households no one was home. In others, the person was willing to be interviewed but was unavailable for any of a number of reasons (illness, presence required elsewhere, hosting visitors). This variable is a tally of these types of non-response.

A.8.3 Calculation of Weights

Enumeration Area Weights

Let $i = 1, \dots, 602$ represent enumeration area i within our sample. Recall that the urban enumeration areas were stratified by region (Eastern, Northern, Southern, Western) and size (<1000, 1000-1999, 2000-2999, 3000-4999, 5000-9999, >10000), and the rural enumeration areas were stratified by chiefdom. First the basic enumeration area weight (EAW) is created using the inverse of the probability of selection within the strata. Enumeration area weights for urban areas range from 2.5 to 4; with a median of 2.88 and a mean of 2.91. Enumeration area weights for rural areas range from 1 to 32; with a median of 4 and a mean of 5.17. Because we used estimated population size to determine both the strata for the urban areas and the number of EAS per stratum, the enumeration area weights are noticeably different for the rural areas than for the urban areas. The EA weights are similar across all sampled urban EAS. In the case of the rural weights, however, we included the condition that at least one EA needed to be sampled from every chiefdom, regardless of underlying population size. For this reason, different sampled rural EAS can represent quite different numbers of total rural EAS.

Next, a raking factor (EARF) is applied to the urban EA weights to adjust them to be valid at the district level. Recall that urban enumeration areas were not sampled within district; as a result, the sum of the urban enumeration area weights may not equal the count of the urban enumeration areas in the district in question. This is corrected by multiplying the urban EA weights for a district by the number of urban EAS in that district, divided by their sum. Raking factors range from .76 to 2.53 for the urban enumeration areas (they are 0.76, .77, .81, .82, .84, .87, .89, 1, 1.03, 1.14, 1.19, 1.21, 2.53), for the rural EAS the raking factor is set to one. The largest raking factor is for Kambia (2.53) and the smallest is for Bombali (.76). The fewer urban EAS there are in a district, the more likely it is that the raking factor will be further away from one; Kambia is notable for having very few urban EAS. Adjusted enumeration area weights for urban areas range from 2.02 to 7.24, with a median of 2.96 and an average of 2.91.

The equation for the final urban enumeration area weights (UEAWs) is as follows:

$$UEAW_i \ll \frac{\#of\ EAs\ in\ stratum\ EA\ i}{\#of\ EAs\ selected\ in\ stratum\ EA\ i} \times \frac{\#of\ EAs\ within\ district\ for\ EA\ i}{\#of\ EAs\ sampled\ within\ district\ for\ EA\ i}$$

For the rural enumeration areas, UEAW is equivalent to EAW.

Village Weights

Village area weights (VWEAWs) are the inverse of the probability that a village was selected, determined by dividing the total number of villages in the enumeration area by the number of villages sampled within the enumeration

area.⁷⁷ Although the counts of villages in rural EAS are based on the 1980 EA maps, we used two procedures to update them. Some of the maps were divided into grids using latitude and longitude lines. The interviewers randomly picked a grid and then went to that grid to see if there was a new village there. If not, they would move to the next randomly selected grid. This helped us account for new villages. Second, upon entering a new chiefdom field staff were instructed to visit the chief and ask him or another official to update the maps. This helped us to discover both new villages and abandoned villages and mark them appropriately on the maps.

Let $i = 1, \dots, 407$ represent rural enumeration area i within our sample. The equation for the final village weights (VWOs) for the rural EAs is as follows:

$$VWO_i = UEAW_i \times VWEAW_i$$

For the urban enumeration areas, VWO is equivalent to UEAW.

Household weights

The initial household weights (HHWOs) are the final village weights (VWOs) multiplied by the final household/housing unit count (FHC) and divided by the number of households/buildings sampled in the village/EA (HUC). Details on how FHC and HUC are formed appear in the variable descriptions above.

Once the HHWOs are formed, two adjustment factors are applied to them: the multiple household adjustment factor (MHAF) and the non-household building adjustment (NHBA). Note that both of these adjustment factors are set to 1 for villages with a FHC below 50. Explanations for how these two adjustment factors are formed appear above.

Let $ji = 1, \dots, ni$ represent household ji within the urban enumeration area i or village i , and recall that iq indicates an urban enumeration area, or village with FHC of 50 or greater, within the area q (qEF , CF , WF , Eastern, Southern, Northern), and qi indicates an area that contains enumeration area/village i . The household weights now take the following form, and are called the ‘provisional urban household weights’:

$$PUHW_{ji} = UEAW_i \times VWEAW_i \times HHVO_i \times MHAF_{qi} \times NHBA_{qi}$$

Although the name ‘PUHW’ implies that it is a variable for urban EAS only, all EAS urban and rural have a value for PUHW.

After the provisional urban household weights have been determined, they are multiplied by a raking factor (HHRF) formed from the ratio of the sum of the provisional urban household weights for the chiefdom to the 2004 census count of households for that same chiefdom. This way, the sum of the households will equal the number of households reported for that area in 2004.

Let ki represent the chiefdom for urban enumeration area i or village i . Then equation for the final household weight

⁷⁷Although some maps were sectioned into grids and a square in the grid was randomly selected, then one village was randomly selected from the grid, the vast majority of the squares on the grid did not contain more than one village. We therefore choose to ignore any variations in sampling probability caused by the grid system for these calculations. The sum of the village weights for a rural EA equals the total number of villages in the rural EA.

(HW) is then as follows:

$$HW_{ji} = PUHW_{ji} \times HHRF_{ki}$$

Individual Weights

Provisional weights for individual members of the households are formed by multiplying the household weight by the count of household members obtained during the interviews (IWHC). These weights are then multiplied by a raking factor (IRF) that is the ratio of the counts of individuals provided for the chiefdom by the 2004 census to the sum of the provisional individual urban and rural weights for the district or urban area. In other words, the formula for the individual weights (IWs) is:

$$IW_{ji} = HW_{ji} \times (\#HH \text{ members in } HH_{ji}) \frac{\#of \text{ individuals within area/district of } HH_{ji}}{\sum(HW_{ji} \times (\#HH \text{ member in } HH_{ji}))}$$

A.8.4 Multiple Systems Estimation

The survey analysis does not provide an estimate of the deaths due to violence during the civil war. As discussed in the text, the three samples make an analysis only of reported deaths. Since each of the documentation projects was itself limited not all Sierra Leoneans were interviewed the reported deaths necessarily omit many deaths due to violence which occurred during the war. Earlier, unscientific estimates of the total deaths due to killing during the war ranged up to 50,000; with approximately 5,000 documented deaths, there are clearly many more killings which have not yet been reported.

Multiple systems estimation (MSE) is a statistical technique that uses several separately-collected incomplete lists of the population. The lists are linked by identifying the elements common across lists in order to estimate the number of elements that are missing from all of the lists. In this project, killings documented in the datasets collected by the TRC, the CGG, and the SLWCD were matched across the three systems using the blocking and linking techniques described above.⁷⁸ The most basic form of this technique is capture-tag-recapture, which uses only two lists. The mathematical basis of the three-list method can be derived from this.⁷⁹

A technical explanation of how a count of the unknown members of the population can be estimated using capture-tag-recapture is as follows. Consider the case of two projects P1 (a list of A individuals) and P2 (a list of B individuals). There are M individuals who are linked across both lists, in a universe of N total individuals (N is unknown). If all of the people in the universe N have an equal probability of appearing in List 1, then the probability of a specific

⁷⁸This explanation follows P. Ball, J. Asher, D. Sulmont, D. Manrique, '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, a report to the Peruvian Truth and Reconciliation Commission.' Washington, DC: AAAS. 28 August 2004. Available online at <http://shr.aaas.org/hrdag/peru>. Also see Romesh Silva and Patrick Ball, '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, Section 5.7. 9 February 2006. Available online at <http://www.hrdag.org/timor> (2006-03-07).

⁷⁹The initial application of multiple-systems estimation to demographic estimation was by C. Chandra Sekar and W. Edwards Deming, 'On a Method of Estimating Birth and Death Rates and the Extent of Registration,' *Journal of the American Statistical Association*, March 1949: 101-115. A thorough discussion of the estimators for the dual-system approach and the relevant error calculations is available in Yvonne M. M. Bishop, Stephen E. Fienberg, and Paul H. Holland. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press. 1975. For commentary on the use of these methods in human rights analysis, see Fritz Scheuren, 'History Corner,' *The American Statistician*, February 2004.

individual being reported by P1 is

$$\Pr(\text{captured in list 1}) = (A / N)$$

Similarly, if all of the people in universe N have an equal probability of appearing in List 2, then the probability of a specific individual being reported by P2 is

$$\Pr(\text{captured in list 2}) = (B / N)$$

The probability of a specific individual being captured in both lists is

$$\Pr(\text{captured in list 1 and list 2}) = (M / N)$$

By definition, the probability of an event composed of two independent events is the product of the independent probabilities. Therefore,

$$\Pr(\text{captured in lists 1 and 2}) = \Pr(\text{captured in list 1}) \times \Pr(\text{captured in list 2})$$

Which is $(M / N) = (A / N) \times (B / N)$: given this equation, solve for N. Rearranging the terms, $(M / N) = (A \times B) / (N \times N)$ and then multiplying by N, $M = (A \times B) / N$ multiplying again $M \times N = A \times B$, and finally dividing by M yields $N = (A \times B) / M$. Note that with the final equation, the total number of deaths N can be estimated using the totals from A and B and from the links between them, M.

There are three principal assumptions implicit in this solution. The first assumption is that none of the lists has internal duplication (i.e., that individuals reported more than once have been identified and controlled), and that the linkage between the lists is accurate. In this project these two assumptions were controlled during the automated record linkage as described above. In additional research to follow from this project, we will conduct extensive simulations to test the effect of different assumptions about how Sierra Leoneans identified people who had died on the total estimated number of deaths.

The second assumption assumes that individuals are not entering or leaving the universe during the process of creating the lists, and second that the lists were selected randomly from the population. In human rights documentation projects, the first part of this assumption is irrelevant because the documentation occurs retrospectively. The second assumption cannot be satisfied (except in the case of a probability survey, such as the SLWCD or the Retrospective Mortality Survey in Timor-Leste), and it must be replaced by the assumption that the estimation is robust to the selection process. This assumption is managed by making the estimates at the finest possible level of stratification, as described below.

Another assumption is that the lists are independent, that is that the probability that an individual is in list two (P2) is independent of the probability that the individual is captured in list one (P1). In this project, we have three lists available (SLWCD, CGG, and TRC), and so it is possible to model the interdependencies or correlations in capture probabilities among the lists. The results of the modeling are described at the end of this section.

Autostratification

This project used a new technique called ‘autostratification.’ In order to analyze patterns, it is useful to have estimates for each type of violation, for each year, and for each district, for example. However, dividing the data into such a small fraction (called a ‘stratum’) may mean that there is insufficient data to make an estimate using MSE.

When a stratum is too small, it can be added to an adjacent stratum, such as a neighboring district or the previous or following year. Note that a stratum that is too small cannot be added to an arbitrary stratum, but instead must be combined with a stratum that logically connects to it. The balance is to create strata that are as small as possible without being too small. It is difficult to find these optimally balanced strata by hand, as the analyst must painstakingly combine and recombine different groups of space and time in order to find blocks that have just the right amount of data.

For this project, we created software to perform autostratification. We defined very fine strata (districts by year for killings). The software then tried to make an MSE estimate for each stratum. For the strata which could not be estimated, the software then tried to add the stratum to each of its neighbors, in turn; if it could not find an estimate, it tried combinations of neighbors, and then each neighbor with that neighbor’s neighbors; and so forth, permitting the aggregated strata include up to 7 individual strata.

There were 137 strata crossing the 13 districts with the years between 1991-2002 (some year-district combinations had no data and could not be included). MSE was attempted using various stratification techniques and varying assumptions about the appropriate strictness of the record linkage.

Results of MSE

When estimating MSE, there are seven possible models for each block of data which are to be analyzed. Our technique estimates all seven models for each stratum, and chooses the model which minimizes the Bayesian Information Criterion (BIC). This statistic measures the ‘relative plausibility’ of a model capable of making an estimate relative to a model that uses all the available information and therefore fits the data perfectly (a ‘saturated’ model). The model with the lowest BIC for a given set of data will be the model that has produced the most reliable estimate using the least complicated model. If the BIC is less than zero, the model can be said to ‘fit’ the data.⁸⁰ Our practice is to choose the model for each stratum which has the minimum BIC; if no model has a BIC \leq 0, with occasional exceptions, the stratum cannot be estimated.

With the current status of development, models can be estimated for varying levels of strictness in the record linkage: stricter record-linkage means that records must be more closely related before a match will be assigned, and less strict record-linkage means that records with more varying values will be accepted as referring to the same person. Thus stricter matching means that more records will be identified as the same person, and consequently fewer unique individuals will be identified. The strictness is measured by the link weight, defined above.

Note that varying the weights used in record-linkage varies the resulting estimate by a factor of 3, from 10,000 to

⁸⁰See Daniel Powers and Yu Xie, *Statistical Methods for Categorical Data Analysis*, Academic Press. 2000. In practice, a BIC of approximately zero corresponds to a chi-square measure (with the appropriate degrees of freedom) significant at the 0.05 level.

30,000. From our analysis of the records linked by the software, these weights define the range of plausible values for the matching strictness. We cannot yet narrow the estimate more closely than this range. We expect that with additional research, we will be able to more reliably and scientifically establish the correct values of the weight and other parameters of the record linkage process, thereby creating a replicable and transparent process for record linkage and MSE.

B Acknowledgments

Tamy Guberek, Daniel Guzmán, Patrick Ball and Romesh Silva designed and conducted the statistical analysis. Scott Weikart wrote the code to clean the datasets, standardize names, de-duplicate datasets and cross-link datasets. Rapheal Kaplan wrote the Analyzer databases for all three databases, managed the data flow among varying formats, and, with Patrick Ball, wrote some of the algorithms needed for the cross-linking of datasets. Wendy M. Grossman wrote the text of the report. Kristen Cibelli assisted with background research, logistical planning for the SLWCD Survey implementation and training of the survey teams in Sierra Leone. Jana Dudukovic canonicalized the individual names of victims in the databases and performed quality control on the data input. Jana Asher of SHR-AAAS designed the sampling plan and questionnaire for the SLWCD Survey, managed the survey implementation in Sierra Leone, developed the weights for the survey-based estimates, calculated the survey-based estimates, and wrote the section on the survey methodology and estimation. Richard Conibere managed the data coding and data entry teams for all three databases and also conducted preliminary data preparation of the raw data from the SLWCD Survey.

This project was funded in part by a subcontract from the American Bar Association with funds from the United States Department of State. The authors thank the John D. and Catherine T. MacArthur Foundation, The Oak Foundation, and the Packard Humanities Institute, all of whom provide core support to the Benetech Human Rights Program that was used to complete this project.

This publication was made possible through the support provided by the United States Department of State Bureau for Democracy, Human Rights and Labor, under the terms of the grant no. S-LMAQM-04-GR-026. The opinions expressed here in are those of the authors and do not necessarily reflect the view of the United States Department of State.

The materials contained herein represent the opinions of the authors and editors and should not be construed to be the view of the Benetech Initiative, the American Bar Association (ABA), the Science and Human Rights Program of the American Association for the Advancement of Science (SHR-AAAS), or any of the other contributing organizations. The interpretations and conclusions are those of the authors and do not purport to represent the views of the Benetech Board of Directors, any of Benetech's constituent projects, or the donors to Benetech. Likewise, the views expressed herein have not been approved by the House of Delegates or the Board of Governors of the ABA and, accordingly should not be construed as representing the policy of the ABA. The AAAS Committee on Scientific Freedom and Responsibility (CSFR), in accordance with its mandate and Association policy, supports publication of this report as a scientific contribution to human rights; however, interpretations and conclusions are those of the authors and do not purport to represent the views of the AAAS Board, Council, the CSFR, or the members of the Association.

This report should be cited as:

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.

Certain rights are granted under the Creative Commons Attribution-NonCommercial-ShareAlike license, available on the web at: <http://creativecommons.org/licenses/by-nc-sa/1.0/legalcode>

The license terms are summarized here:

Attribution: The licensor permits others to copy, distribute, display, and perform the work. In return, licensees must give the original author credit.

Noncommercial: The licensor permits others to copy, distribute, display, and perform the work. In return, licensees may not use the work for commercial purposes, unless they get the licensor's permission.

Share Alike: The licensor permits others to distribute derivative works only under a license identical to the one that governs the licensor's work.

Contact Information:

The Benetech Initiative

tel: +1 650-475-5440

fax: +1 650-475-1066

Email: info@benetech.org

Web: <http://www.benetech.org>