

---

# Metodología intermuestra

DOCUMENTO ELABORADO POR LA AAAS<sup>1</sup>

## Introducción:

### *El objetivo y el problema*

**1.** ¿Cuántas personas fueron asesinadas en Guatemala durante el período del mandato de la CEH, 1960-1996? Tomando en cuenta la información de las tres bases de datos sobre violaciones de Derechos Humanos (CEH, CIIDH, REMHI), recogida por medio de entrevistas con testigos directos y víctimas, surgen algunas preguntas: a) ¿cuántos asesinatos fueron documentados por los tres proyectos? y b) ¿cuántos asesinatos se estiman que ocurrieron —en total— durante el conflicto armado?

**2.** Este análisis trata los dos temas. En la parte I, se establece la información recogida por los tres proyectos y los objetivos del análisis. En la parte II se explican los métodos científicos utilizados para estimar cifras que responden las preguntas empíricas mencionadas. En la parte III se presentan los resultados con interpretaciones. La parte IV, se enfoca en cuatro regiones en las que, posiblemente, ocurrió genocidio en el período 1981-1983 y se presenta los resultados desagregados según la etnia de las víctimas. Finalmente, en la parte V se presentan las diversas formas en que los tres proyectos concuerdan en términos estadísticos, sobre los patrones y tendencias de violencia en Guatemala.

## **I. La información recogida**

**3.** Las tres bases de datos que han sido elaboradas con información recogida en entrevistas con testigos y víctimas, contienen listas de víctimas de asesinato obtenidas de cada proyecto. No obstante, no tenían definiciones precisas en común. La medida en este estudio es *muertos*, y no la categoría más jurídicamente precisa de “ejecución extrajudicial” utilizada en el informe de la CEH. Las tres proyectos tenían definiciones distintas de asesina-

to, y para juntalos, fue necesario usar la definición del fenómeno más amplio. Las cifras de la CEH en este estudio deben ser comparados con los de totales de muertos en el informe.

4. Muchos asesinatos no fueron reportados a ningún proyecto y, por eso, la cantidad expuesta en la Tabla 1 es menor que el total de víctimas que ocurrieron en la realidad. El total, en cada una de las tres bases se presenta, por región y por época, en la última fila de la Tabla 1.

**TABLA 1.<sup>2</sup>**  
**Cantidad de asesinatos documentados**  
**(víctimas con y sin nombres), por época, región, y proyecto**

Región	CEH		CIIDH <sup>3</sup>		REMHI	
	1960-77	1978-96	1960-77	1978-96	1960-77	1978-96
Región 0 (otros)	271	9,916	18	2,586	84	6,888
Región I (área Ixil)	14	4,609	0	4,028	9	5,423
Región II (Cahabón)	1	532	0	135	7	453
Región III (Rabinal)	0	1,379	0	1,297	0	1,354
Región IV (San Martín Jilo.)	0	1,347	1	20	0	68
Región V (Nte. De Huehue)	0	1,746	0	1	0	1,032
Región VI (área Zacualpa) <sup>4</sup>	0	1,951	0	238	1	1,674
Región VII (Guatemala)	91	245	1	15	10	111
Región VIII (área Panzós)	0	169	11	41	1	1,167
Región IX (Ixcán)	3	2,421	0	48	5	2,751
Región X (área Costa Sur)	25	190	2	91	23	139
Total	405	24,505	33	8,500	140	21,060

5. Nótese que los tres proyectos no cubrieron todas las regiones: los tres realizaron investigaciones en el área ixil sobre violaciones en la época 1978-1996, pero solamente la CEH recogió información adecuada en San Martín Jilotepeque.<sup>5</sup> También es claro que ninguno de los tres proyectos cubrió bien la época 1960-1977, que incluye las matanzas de los años 1968-1973. Cualquier estimación tiene que tomar en cuenta estas limitaciones.<sup>6</sup>

6. Si no hubiese víctimas reportadas en más de un proyecto, el total de víctimas documentadas sería la suma de las tres bases de datos, es decir, la suma de la última línea de la tabla,  $405+24,505+33+8,500+140+21,060 = 54,643$ . Sin embargo, muchas víctimas fueron reportadas en dos o tres proyectos, por eso, no se puede presumir que el total de víctimas documentadas sea igual a la suma simple. Los totales de cada proyecto no se pueden sumar para estimar el total de asesinados reportados.

7. Los proyectos fueron manejados independientemente: por eso la misma víctima puede estar reportada en más de un proyecto. Por ejemplo, imagine que Juan Pérez fue asesinado. La esposa de Juan Pérez pudo haber reportado su asesinato al CIIDH en 1994; el hijo del mismo señor Pérez

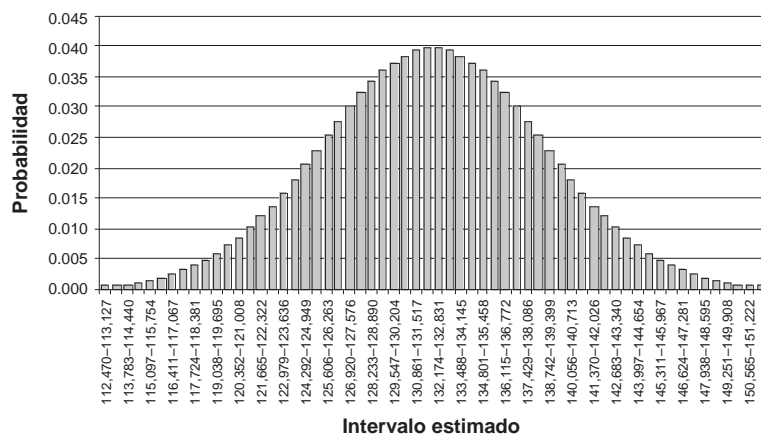
pudo dar su testimonio a REMHI en 1996; y el vecino del señor Pérez pudo relatar el mismo hecho a la CEH en 1997. Si se sumaran las tres bases de datos, la violación que pasó a Juan Pérez sería contabilizada tres veces.

**8.** El reportaje duplicado de violaciones en más de una base de datos es la causa del fenómeno de “traslape” entre bases de datos. Para estimar el número total de víctimas, los traslapes entre las bases de datos tienen que ser estimados y las sumas finales reducidas en esa cantidad.

**9.** Hay dos posibilidades que muestran los límites del problema de traslape. Como sugerimos antes, si no hubiese ninguna víctima que aparezca en más de una base de datos, la suma lógica y correcta de las tres sería igual a la suma aritmética de las tres (54,643); éste se llama el límite superior. El otro límite, el inferior, representa la posibilidad de que la mayor de las tres bases (la CEH) contenga todos los casos de los otros. En esta situación, la cantidad total de asesinados sería igual al número reportado por la más grande, la CEH ( $405+24,505=24,910$ ). La cantidad de víctimas únicas sumadas, de las tres bases de datos, cae entre estos límites, es decir, entre 24,910 y 54,643. El análisis para estimar la cantidad de asesinatos documentados es para reducir este rango.

**10.** También resulta que muchos asesinatos nunca fueron reportados a ningún proyecto. En la sección de abajo, se expone una estimación de la cantidad de asesinatos<sup>6</sup> no considerados por la CEH, el CIIDH, ni el REMHI. La estimación de este análisis es que ocurrieron entre 119,300–145,000 asesinatos, siendo la máxima probabilidad de 132,000. La distribución de probabilidades alrededor de la estimación de 132,000 se presenta en la Gráfica 1. Esta cifra se justifica en las siguientes secciones.

**GRÁFICA 1:**  
**Probabilidad de que la cifra real caiga dentro del intervalo**



## II. Sumar víctimas de asesinato sobre tres bases de datos independientes

### *Análisis de “traslape”*

**11.** La información contenida en las tres bases de datos representa víctimas identificadas o enumeradas por los testigos; algunas (pero no todas) las víctimas fueron identificadas por nombre.<sup>7</sup> La cantidad de asesinados en cada base está identificada por la notación expuesta abajo.

$M_{CEH}$  = la cantidad de víctimas en la base de datos de la CEH

$M_{CIIDH}$  = la cantidad de víctimas en la base de datos del CIIDH

$M_{REMHI}$  = la cantidad de víctimas en la base de datos de la REMHI

**12.** Ninguna de las bases de datos estimó, en forma directa, cuántos asesinados ocurrieron en total en todo el país. Cada base es un listado de las víctimas de matanza que fueron reportadas directamente al proyecto y verificado según la metodología del mismo proyecto. Como ya fue mencionado, muchas víctimas no fueron reportadas a ninguno de los tres proyectos. El total de las víctimas asesinadas en Guatemala puede estar representado en las ocho categorías expuestas abajo.

$N_{000}$  = víctimas que no fueron reportados a ninguna de las tres: la CEH, el CIIDH, ni el REMHI

$N_{111}$  = víctimas que fueron reportados a todas (los tres proyectos)

$N_{110}$  = víctimas reportadas a la CEH y al CIIDH, pero no al REMHI

$N_{101}$  = víctimas reportadas a la CEH y al REMHI, pero no al CIIDH

$N_{011}$  = víctimas reportadas al CIIDH y al REMHI, pero no a la CEH

$N_{100}$  = víctimas reportadas solamente a la CEH, y no al CIIDH ni al REMHI

$N_{010}$  = víctimas reportadas solamente al CIIDH, y no a la CEH ni al REMHI

$N_{001}$  = víctimas reportadas solamente al REMHI, y no a la CEH ni al CIIDH

**13.** El total de víctimas de asesinato en Guatemala,  $N$ , es la suma de estas ocho categorías. El total de víctimas reportadas a uno, dos o tres proyectos,  $N_k$ , es la suma de las siete categorías que son calculables directamente de las bases de datos, es decir,  $N_{111}$  a  $N_{001}$ . Véase la ecuación 1.

**Ecuación 1** 
$$\hat{N}_k = N_{111} + N_{110} + N_{101} + N_{011} + N_{100} + N_{010} + N_{001}$$

**14.** Para contabilizar la cantidad de víctimas reportadas a uno o más proyectos, hay que estimar el número de víctimas reportado a todos los proyectos ( $N_{111}$ , concordado tres veces) y entre cada pareja de proyectos ( $N_{110}$ ,  $N_{101}$ , y  $N_{011}$ ). El complemento de cada proyecto ( $N_{100}$ ,  $N_{010}$ , and  $N_{001}$ ) determina las cifras requeridas para  $N_k$ .

### *Concordancia*

**15.** Es muy difícil encontrar la misma víctima en dos bases de datos con un programa de computadora. Las víctimas son reportadas a los tres proyectos, con información que varía en cada reporte. Los nombres y apellidos de las mismas personas son deletreados de distinta forma, por ejemplo, por ser traducido de idiomas mayas a español en forma inconsistente, los reportes de fechas de nacimiento y de defunción pueden estar equivocados.

**16.** No es práctico ni preciso concordar bases en forma automatizada, con programas de computadora. Para encontrar a una persona de una base de datos (la fuente) en otra de las bases (el objeto), un analista tiene que comparar la persona de la fuente contra todos los registros en el objeto. Este proceso se llama concordar a las bases.<sup>8</sup>

**17.** Concordar las bases es tedioso y requiere mucho tiempo. A cambio de intentar de concordar todos los registros de cada base de datos contra las otras, se tomaron muestras aleatorias de las víctimas identificadas por nombre de cada una de las tres bases de datos.<sup>9</sup> El número de registros tomado en cada muestra está denotada por la letra  $m$  ( $m_{CEH}$ ,  $m_{CHDH}$ ,  $m_{REMHI}$ ). Incluyendo todas las regiones, la cantidad de registros seleccionados de los tres proyectos fue de 1,412, 1,351, y 1,122, respectivamente (Tabla 2).<sup>10</sup>

**18.** Cada persona, en la muestra de cada base, fue comparada con todos los registros en las demás bases. Cuando la misma persona fue encontrada en una de las demás bases, fue anotada como concordada doble; cuando la misma persona fue encontrada en las tres bases, fue anotada como concordada triple.

**19.** Se tomaron cuatro grupos de muestras de las tres bases; un analista trató tres grupos, y el otro trató uno. Solamente se encontraron diferencias muy pequeñas entre las tasas de concordancia que resultaron del trabajo de los dos analistas; la implicación es que el error que resultó de factores fuera del proceso de tomar la muestra fueron mínimos.<sup>11</sup> Los números de personas concordadas, tomando todas las regiones, se presentan en la Tabla 2.

**TABLA 2:**  
**Cantidad de concordados encontrados entre las tres bases**

	CEH	CIIDH	REMHI
$m_{111}$	21	73	19
$m_{110}$	48	153	
$m_{101}$	210		226
$m_{011}$		121	27
$m_{100}$	1,133		
$m_{010}$		1,004	
$m_{001}$			850
Muestra total	1,412	1,351	1,122

**20.** La Tabla 2 indica que en la muestra de 1,412 víctimas seleccionadas de la base de datos de la CEH, 21 fueron encontradas en la base de datos del CIIDH y también en la base de datos de REMHI; esos 21 son triples. También, se encontraron 48 registros de la CEH en la base de datos del CIIDH pero no en REMHI, y 210 más de la CEH fueron encontrados en la base de datos del REMHI pero no en la del CIIDH; 1,133 de los registros seleccionados de la base de datos de la CEH no fueron encontrados en ninguna de las otras dos.

**21.** Al dividir cada  $m_{xyz}$  (Tabla 2), por el total de la muestra de cada base, los números en Tabla 2 se convierten a tasas. Las tasas de concordancia se presentan en la Tabla 3.

**TABLA 3:**  
**Tasas de concordancia para las tres bases de datos**

	CEH	CIIDH	REMHI
$r_{111}$	1.5%	5.4%	1.7%
$r_{110}$	3.4%	11.3%	
$r_{101}$	14.9%		20.1%
$r_{011}$		9.0%	2.4%
$r_{100}$	80.2%		
$r_{010}$		74.3%	
$r_{001}$			75.8%

**22.** En la segunda línea  $r_{110}$  indica que 3.4% de las víctimas en la base de datos de la CEH están también en la base de datos del CIIDH. Pero la base de datos del CIIDH es más pequeña que la base de la CEH, y por eso, la misma estimación, desde el punto de vista del CIIDH, es que 11.3% de las víctimas registradas en la base de datos del CIIDH están en la base de datos de la CEH.

**23.** Nótese que las diferencias en las estimaciones de las tasas no están exactamente en proporción a las diferencias de tamaño entre las bases. Estas diferencias entran por la variación que resultó del proceso de tomar la muestra aleatoria, y por el error del proceso de concordar las bases, y está tratada en la sección de análisis de error (abajo).

### *Estimación de totales de víctimas reportadas*

Como se ha mencionado en la sección de Análisis, el número de asesinatos se estima por la suma de los siete términos de “traslape”, como se presenta en la Ecuación 1.

$$\hat{N}_k = N_{111} + N_{110} + N_{101} + N_{011} + N_{100} + N_{010} + N_{001}$$

**24.** Si se aplican las tasas a la Tabla 3, arriba, a la última línea de la Tabla 1 (total de víctimas 1978-1996), se saca una estimación de los componentes de  $N_k$  basada en la información de las tres bases de datos. Véase el resultado en Tabla 4.

**TABLA 4:**  
**Cantidad de asesinatos, estimados por categoría y proyecto**

	CEH	CIIDH	REMHI
$N_{111}$	364	459	357
$N_{110}$	833	963	
$N_{101}$	3,645		4,242
$N_{011}$		761	507
$N_{100}$	19,663		
$N_{010}$		6,317	
$N_{001}$			15,955

**25.** Sin embargo, para estimar un  $N_k$  en común entre las tres bases de datos, hay que reconciliar las varias estimaciones del número de registros que concuerdan entre las bases de datos ( $N_{111}$ ,  $N_{110}$ ,  $N_{101}$ , and  $N_{011}$ ). Se toma el promedio de los valores en cada fila en la Tabla 4, en que hay más de una estimación, y así se estima la suma en Tabla 5.

Así, el número estimado de asesinatos reportados a los tres proyectos sin duplicación es aproximadamente 47,803. Pero como será mostrado abajo, este número está sujeto a algunos sesgos que se pueden controlar.

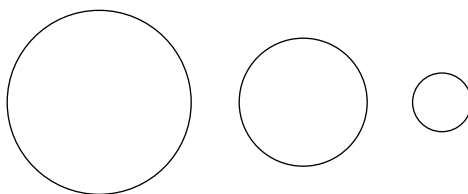
**TABLA 5:**  
**Número estimado de asesinados en las tres bases de datos**  
**(CEH, CIIDH, REMHI)**

	Promedio
$N_{111}$	393
$N_{110}$	898
$N_{101}$	3,943
$N_{011}$	634
$N_{100}$	19,663
$N_{010}$	6317
$N_{001}$	15,955
$N_k$	47,803

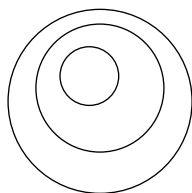
### *Teoría demográfica y la estimación de $N_{000}$*

En el apartado anterior se explica cómo estimar el número de asesinatos documentados. Todavía está pendiente una estimación del número de asesinatos no documentados,  $N_{000}$  en la notación definida. Abajo, en las Gráficas 2a, 2b, y 2c, se presentan tres posibilidades de cómo están relacionados los proyectos.

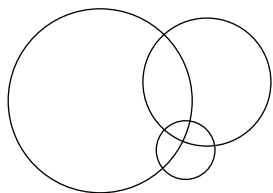
**GRÁFICA 2A:**  
**Distribución de proyectos en el universo**  
**de violaciones (ninguna relación)**



**GRÁFICA 2B:**  
**Distribución de proyectos en el universo**  
**de violaciones (equivalencia completa)**



**GRÁFICA 2C:**  
**Distribución de proyectos en el universo**  
**de violaciones (duplicación parcial)**



En la Gráfica 2a, los proyectos no comparten ninguna violación. En Gráfica 2b, todas las violaciones están contenidas en el proyecto más grande. En Gráfica 2c, algunas violaciones están compartidas. Del apartado anterior, se infiere que Gráfica 2c representa la realidad de los tres proyectos.

Presumamos, por el momento, que los proyectos son independientes, es decir, que la probabilidad de que un testimonio a un proyecto tiene correlación cero con la probabilidad de que el mismo declarante dé el testimonio a otros proyectos. ¿Qué implicación sobre el universo tiene cada Gráfica? En Gráfica 2a, la implicación es que el universo es inmenso en relación a los proyectos porque trabajando independientemente, los proyectos se distribuyeron en las violaciones sin encontrarse con otros proyectos. En Gráfica 2b, la implicación es a la inversa, que el universo es muy reducido, y los proyectos son equivalentes. Pero en la Gráfica 2c, se ve que los niveles de “traslape” son parciales. Con los niveles de “traslape”, se puede inferir la escala del universo.

Considerar el caso con dos proyectos, A y B, que tiene un “traslape” M en un universo de violaciones N.<sup>12</sup> Nótese que la probabilidad que cualquier

asesinato sea seleccionada por A es  $\Pr(A) = \frac{A}{N}$  es decir que  $N = \frac{A}{\Pr(A)}$

y la probabilidad de que cualquier asesinato sea seleccionada por B es  $\Pr(B) = B/N$ . La probabilidad de un asesinato contenido en las dos bases de datos,  $\Pr(M)$ , es igual a  $\Pr(M) = M/N$ , y por la definición de un evento compuesto de dos eventos independientes,  $\Pr(M) = \Pr(A|B) = \Pr(A) * \Pr(B)$ .

Intercambiando los términos,  $\Pr(A) = \frac{\Pr(M)}{\Pr(B)}$ ,

que reduce a  $\Pr(A) = \frac{M / N}{B / N} = \frac{M}{B}$

Dada la primera relación  $\Pr(A) = \frac{A}{N}$  con el resultado anterior nos da

$$\frac{A}{N} = \frac{M}{B}, \text{ y entonces } N = \frac{AB}{M}. \text{ Para estimar solamente los asesinatos}$$

que fueron excluidos de los dos proyectos,  $N_{00} = \frac{(A-M)(B-M)}{M}$ , o

en la notación de los tres proyectos,

$$\text{Ecuación 2 } N_{00} = \frac{N_{10} * N_{01}}{N_{11}}$$

Con la misma lógica, se puede derivar un estimador para  $n_{000}$ : la medida del número de asesinatos que no fueron documentados por ninguno de los tres proyectos.<sup>13</sup> Este estimador está presentado en la Ecuación 3.

$$\text{Ecuación 3 } n_{000} = \frac{n_{100}n_{010} + n_{100}n_{001} + n_{010}n_{001}}{n_{110} + n_{101} + n_{011}}$$

### *Medir error del estimador y de la muestra*<sup>14</sup>

**26.** Los estimadores de  $n_{000}$  y el número total de asesinatos  $\hat{n}$  dependen de los niveles de “traslape” entre las tres bases de datos. Estos niveles (como los presentados en Tabla 3, arriba) se multiplican por el total de violaciones en cada base, para sacar estimaciones para las categorías  $n_{111}$ ,  $n_{110}$ , ...,  $n_{001}$  que suman a  $N_k$ . Con la Ecuación 3, se estima  $n_{000}$ ; sumando, se puede estimar  $\hat{n}$ . No obstante, se requiere otro método que se llama el método de “jackknife” para lograr dos objetivos: a) quitar el sesgo de razón que resulta de la Ecuación 3; y b) para determinar la variación de los tres estimadores necesarios para este análisis ( $N_k$ ,  $n_{000}$ , y  $\hat{n}$ ). En la explicación general del método (abajo), el estimador  $\theta$  representa a cada uno de los tres en los cálculos. Por ejemplo,  $N_k$  en la Tabla 5 (arriba), 47,803, es  $\theta$  para  $N_k$  tomada a nivel nacional.

**27.** El método empieza por dividir aleatoriamente la muestra de los registros concordados (que tiene un total de  $n$  registros) a  $k$  grupos, conteniendo cada cual  $m$  registros:  $n = mk$ .  $\theta_{(\alpha)}$  está calculada de la misma manera como  $\theta$  pero con una muestra reducida de tamaño  $m(k-1)$  obtenida por omitir el grupo  $\alpha$ . Definir

$$\text{Ecuación 4 } \hat{q}_a = k\hat{q} - (k-1)\hat{q}_{(a)}$$

$$\text{Ecuación 5 } \hat{q} = \frac{1}{k} \sum_{a=1}^k \hat{q}_a$$

**28.** La Ecuación 4 nos da un vector de tamaño  $k$  de calculado de  $\mathcal{J}_v$  las muestras reducidas por omitir los grupos; el promedio de vector de  $\mathcal{J}_v$  es  $\hat{\pi}$  (véase Ecuación 5), se llama el “estimador de Quenouille”, que quita varios sesgos que afectan a  $\theta$ . Este estimador es lo que se presenta en la Tabla 7 (Resultados, en la Parte III).

**29.** El otro resultado beneficioso del método de “jackknife” es si los valores de  $\theta_{(a)}$  están distribuidos normalmente.<sup>15</sup> El error estándar del estimador (la raíz cuadrada de la variación) se estima con la Ecuación 6.

$$\text{Ecuación 6 } SE(\hat{q}) = \sqrt{\frac{1}{k(k-1)} \sum_{a=1}^k (\hat{q}_a - \hat{q})^2}$$

**30.** Los errores estandares presentados en la Tabla 7 fueron calculados según la Ecuación 6.

### *Cobertura en tiempo y espacio y su efecto en $n_{000}$*

Fue anotado en la discusión de la Tabla 1 que ninguno de los proyectos cubrieron bien las violaciones del período 1960-1977, y por eso ninguna estimación para este período es posible con los datos disponibles. La complicación más importante para la estimación de  $n_{000}$  es que los proyectos no cubrieron todas las regiones con la misma intensidad. Si no se tratan aparte las regiones con niveles de cobertura distintas, se puede sesgar la estimación.

Considerar la región IV, en que la CEH encontró más de 1,300 asesinatos, mientras los otros dos proyectos reportaron solamente algunas docenas. Por supuesto, los niveles de traslape son bajos, pero estos niveles no se pueden utilizar en una estimación de  $n_{000}$  porque el concepto de traslape requiere que los proyectos hubieran trabajado en las mismas áreas. Así que la estimación de  $n_{000}$  tuvo que basarse solamente en aquellos proyectos que realmente lograron trabajar en cada región. Los proyectos que contribuyeron a la estimación de  $n_{000}$  están en la Tabla 6.

**TABLA 6:**  
**Proyectos utilizados para estimar  $n_{000}$ , por región**

Región	Proyectos con cobertura adecuada	Ecuación para $n_{000}$
Región 0 (otro)	Los tres	Ecuación 3
Región I (área Ixil)	Los tres	Ecuación 3
Región II (Cahabón)	Los tres	Ecuación 3
Región III (Rabinal)	Los tres	Ecuación 3
Región IV (San Martín Jilotepeque)	Solamente CEH	No se puede estimar $n_{000}$
Región V (Norte de Huehuetenango)	CEH & REMHI	Ecuación 2
Región VI (área Zacualpa)	Los tres	Ecuación 3
Región VII (Guatemala)	CEH & REMHI	Ecuación 2
Región VIII (área Panzós)	CEH & REMHI	Ecuación 2
Región IX (Ixcán)	CEH & REMHI	Ecuación 2
Región X (área Costa Sur)	Los tres	Ecuación 3

### III. Resultados y algunas limitaciones de la interpretación

Ahora se pueden presentar resultados por categoría y región.

**TABLA 7:**  
**Totales de asesinatos en Guatemala 1978-1996,**  
**por categoría de estimación y región**

Categoría	Región											Total
	0	I	II	III	IV	V	VI	VII	VIII	IX	X	
$N_{111}$	67	141	15	146	0	0	17	2	0	2	2	391
$N_{110}$	378	406	8	98	5	0	67	3	0	16	2	983
$N_{101}$	1,358	1,010	204	170	13	206	336	24	43	681	13	4059
$N_{011}$	133	419	16	122	0	0	0	0	0	0	0	690
$N_{100}$	8,260	3,187	221	1,028	1,325	1,597	1,642	226	156	1,720	182	19,545
$N_{010}$	2,256	2,708	85	836	16	1	195	15	41	30	91	6,274
$N_{001}$	5,228	3,999	295	926	59	765	1,166	77	1,099	2,054	106	15,773
$N_k$ (sin duplicación)	17,679	11,870	844	3,328	1,418	2,569	3,416	347	1,339	4,501	396	47,706
SE ( $N_k$ )	110	135	24	79	11	44	76	9	26	77	5	228
$N_{000}$	38,856	17,397	466	6,467	0	5,548	5,836	561	2,265	5,052	2,019	84,468
SE ( $N_{000}$ )	3,809	2,045	105	1,152	0	1,826	1,890	350	3,062	995	1,840	6,388
$\hat{N}$	56,535	29,267	1,310	9,795	1,418	8,117	9,252	908	3,604	9,553	2,415	132,174
SE ( $\hat{N}$ )	3,918	2,175	127	1,218	11	1,870	1,964	357	3,087	1,072	1,844	6,568

**31.** En la Tabla 7 se ve que en la región 0 fueron documentados  $N_k = 17,679$  asesinatos entre los tres proyectos. Sobre todas las regiones, fueron documentados únicamente 47,706 asesinatos, siendo ésta la suma de las

estimaciones regionales.<sup>16</sup> El error estándar  $SE(N_k)$  no es la suma simple de las regiones, sino es raíz cuadrada de la suma de los valores regionales ( $i=0, 1, \dots, X$ ) al cuadrado:

$$SE(N_{kl}) = \sqrt{\sum_{i=0}^X SE(N_{ki})^2}$$

**32.** Igualmente, los valores para  $N_{000}$  y  $N$  son los valores regionales sumados, y el error estándar para  $N_{000}$  y  $\hat{n}$  es la raíz cuadrada de la suma de los valores regionales al cuadrado. Así que se estima que fueron aproximadamente 84,468 asesinatos que no fueron reportados en la CEH, el CIIDH, ni en el proyecto REMHI. Sumando  $N_k$  y  $n_{000}$  a  $\hat{n}$ , el análisis llega a la estimación final, que hubo 132,174 asesinatos en Guatemala entre 1978-1996, con un error estándar de 6,568.

### *Correcciones posibles y limitaciones en la interpretación de la Tabla 7*

**33.** Hay cinco fuentes de error que no se puede cuantificar en este análisis, por falta de tiempo, recursos o de datos adecuados. En algunas investigaciones preliminares, el efecto global de estas correcciones es conservador, o mejor dicho, tiende a reducir la estimación de  $\hat{n}$ . La conclusión del apartado es que el efecto acumulado de estos sesgos no es significativo.

### *Correlación entre fuentes*

**34.** La estimación de  $n_{000}$  depende de la presunción de independencia entre las fuentes, es decir que la probabilidad de que cada declarante dé su testimonio a un proyecto tiene correlación cero con la probabilidad que el mismo declarante dé su testimonio a otros proyectos. Es cierto que esta correlación no es cero, sino positiva, mayor que cero, por dos razones.

**35.** *Primero*, porque investigaciones realizadas por psicólogos han mostrado que sobrevivientes de violaciones de derechos humanos que pueden dar testimonios, en condiciones que afirman su humanidad, mejoran su estado psicológico. Es probable entonces que personas que dan testimonios en estas condiciones busquen más oportunidades para darlo, incrementando así el nivel de “traslape”.

**36.** *Segundo*, es conocido que varias organizaciones sociales organizaron sus bases para presentar testimonios a los tres proyectos. Así, los miembros de estas organizaciones, que tienen mayor probabilidad de dar su testimonio más de una vez, también aumentan el nivel de “traslape”. Los dos efectos —que son ciertos— tienen el efecto de sesgar la estimación de  $n_{000}$  para abajo.

### *Errores en concordancia*

**37.** Si los analistas que hicieron la concordancia no encontraron víctimas en común, por casualidad o por datos inadecuados en las fuentes originales, estas omisiones tendrían el efecto de bajar el nivel de traslape estimado y, en consecuencia, sesgar la estimación de  $n_{000}$  para arriba. En investigaciones superficiales (lo único que fue posible con los datos actuales) se encuentran efectos mínimos de este tipo, por lo que se tendría una escala de no más que 12% de la estimación final de  $n_{000}$ , implicando 8% a  $\hat{n}$ . Considerando las demás fuentes de sesgo listadas en este apartado, y reconociendo que los datos para el análisis de este error fueron muy reducidos, se decidió no incluir la cuantificación de este error en el análisis final.

### *Duplicación interna*

**38.** Todos los proyectos que reciben información de fuentes primarias tienen el problema de duplicación, que es el resultado de reportes múltiples del mismo hecho.<sup>17</sup> La duplicación interna tiene la tendencia a aumentar artificialmente la estadística de cuántos asesinados están representados en una base de datos. Los tres proyectos tomaron muchas medidas para depurar las duplicaciones internas, pero todavía quedaron algunas. En un análisis preliminar, no aparecen suficientes víctimas duplicadas para requerir una cuantificación.

### *Tasas de traslape entre clases de víctimas*

**39.** La medida de “traslape” entre las tres bases de datos fue tomada sobre víctimas identificadas por un nombre y un apellido. Sin embargo, muchas víctimas no son identificadas con nombre, como consecuencia de violencia masiva que sobrecargó la capacidad de los testigos para recordar a todas las víctimas. Es posible que los niveles de “traslape” entre víctimas no identificadas por nombre, sea mayor o menor que los niveles de “traslape” medidos entre las víctimas identificadas. Dada la dificultad para detectar “traslape” entre víctimas no identificadas, no es posible cuantificar la diferencia entre niveles de “traslape”.

### *Áreas geográficas excluidas por todos los proyectos*

**40.** En la Tabla 7 fue anotado que en la región IV, no hubo datos adecuados para hacer una estimación de  $n_{000}$ . Dada la experiencia de otras regiones en que la razón entre  $N_k$  y  $n_{000}$  varía entre 0.5 y 2, con un modo y promedio cercana a 0.5, es probable que el valor de  $n_{000}$  para la región IV

sea aproximadamente de 2500. Dado que en otras regiones  $N_k$  ésta compuesto por lo menos por dos proyectos, y lo recogido en la región IV corresponde a la CEH solamente, es posible que la razón entre  $N_k$  y  $n_{000}$  para la región IV sea 0.25 o menos, aumentando la estimación por un factor de dos o más. La falta de certeza sobre este número no es verificable por otros métodos, y por eso, no se incluye en la estimación final.

**41.** La región IV es un ejemplo de un problema mucho más fundamental: que esta metodología solamente sirve para las áreas cubiertas por lo menos por dos de los tres proyectos, aunque sólo sea parcialmente. En áreas trabajadas por uno o ninguno de los tres proyectos, no hay una base para hacer una proyección al total de víctimas excluidas ( $n_{000}$ ); en cambio, en estas situaciones solamente  $N_k$  entra a la estimación. Dado el enfoque común de los tres proyectos en áreas de Guatemala, en que las violaciones en gran escala fueron conocidas por fuentes periodísticas o por organizaciones que trabajaban en dichas áreas, no es muy probable que las áreas excluidas dejaran muchas violaciones afuera del análisis. Sin embargo, cada área que fue excluida de los tres proyectos —pero en las que se cometieron asesinatos— aumentaría la estimación.

#### IV. Estimación de tasas de asesinatos, por grupo étnico y región

**42.** La tasa de asesinatos es la proporción de personas de un grupo definido que son asesinadas, es decir, es igual al número de asesinados dividido por el número en total del grupo. La CEH está interesada en las cifras relativas (proporciones) de personas asesinadas, comparadas entre grupos étnicos, durante el período 1981-1983.

**43.** Seis regiones fueron identificadas como aquellas en que según las fuentes secundarias y evidencia anecdotal, la violencia del Estado fue especialmente concentrada contra gente indígena. Las regiones de interés están en el listado de la Tabla 8, con la población según el censo de 1981, por grupo étnico.

**TABLA 8:**

**Población en seis regiones por grupo étnico, 1981**

	Indígena	No- indígena
Región I: área Ixil	38,902	5,882
Región II: Cahabón	20,706	868
Región III: Rabinal	18,610	4,120
Región IV: San Martín Jilotepeque	31,690	4,876
Región V: norte de Huehuetenango	53,556	11,123
Región VI: Chiché, Zacualpa, Joyabaj	51,105	10,997

**44.** Para calcular la tasa de matanza, hay que estimar el número de víctimas; este procedimiento se hace dos veces (obtendremos dos cifras), primero el total de víctimas documentadas por las tres bases de datos, y segundo la cantidad proyectada según la metodología descrita en Parte II. Se requieren los siguientes pasos:

**A.** El número de asesinados que ocurrieron entre 1981-1983, menos los atribuidos a la URNG, fueron calculados por grupo étnico (indígena, no indígena, etnia desconocida) para cada una de las seis regiones en las tres bases de datos. Este paso es, como en la Tabla 1, arriba, limitada al período 1981-1983 y desagregada por etnia de la víctima.

**B.** Las cantidades de concordados y tasas de “traslape” fueron calculadas para cada una de las seis regiones (lógicamente parecida a las Tablas 2 y 3).<sup>18</sup>

**C.** El número de víctimas fue estimado por grupo étnico de las tasas de “traslape” regional y el número de víctimas en cada una de las bases (semejante a la Tabla 4).

**D.** Se toma el promedio del número de víctimas estimadas entre las tres bases de datos (semejante a la Tabla 5) produciendo así, estimaciones de los totales para cada región.

**E.** El método “jackknife” fue aplicado a cada grupo definido por etnia y región, siguiendo con las ecuaciones 4, 5, y 6, para estimar  $N_k$  y  $\hat{n}$  (y sus errores estandar) para cada etnia en cada región. Los valores de  $N_k$  se presentan en la Tablas 9a y 9b, y los de  $\hat{n}$  en la Tablas 11a y 11b.

**TABLA 9A:**  
**Número de asesinatos documentados ( $N_k$ )**  
**en las tres bases de datos, 1981-1983**

		Región I	Región II	Región III	Región IV	Región V	Región VI
Indígena	$N_k$	1388	340	1071	1012	1020	1126
	SE( $N_k$ )	25.5	13.12	33.71	0.67	17.11	16.7
No-Indígena	$N_k$	32	2	13	16	8	6
	SE( $N_k$ )	0.49	0.07	0.33	0.14	0.12	0.13
Sin etnia conocida	$N_k$	4339	186	1669	10	752	1208
	SE( $N_k$ )	62.87	9.02	47.75	0.64	31.95	46.79

F. Tomando de la Tabla 9a las víctimas con etnia conocida, las víctimas sin etnia conocida fueron proporcionadas a las categoría “indígena” o la categoría “no-indígena” según las proporciones en Tabla 10 a la Tabla 9b.

**TABLA 9B:**  
**Número de asesinatos documentados (N<sub>k</sub>)**  
**en las tres bases de datos, por grupo étnico incluyendo víctimas**  
**sin identificar, 1981-1983**

		Región I	Región II	Región III	Región IV	Región V	Región VI
Indígena	N <sub>k</sub>	5,632	525	2,720	1,022	1,767	2,327
	SE(N <sub>k</sub> )	66.56	15.90	57.98	0.92	36.05	49.42
No-Indígena	N <sub>k</sub>	127	3	33	16	13	13
	SE(N <sub>k</sub> )	0.49	0.07	0.33	0.14	0.12	0.13

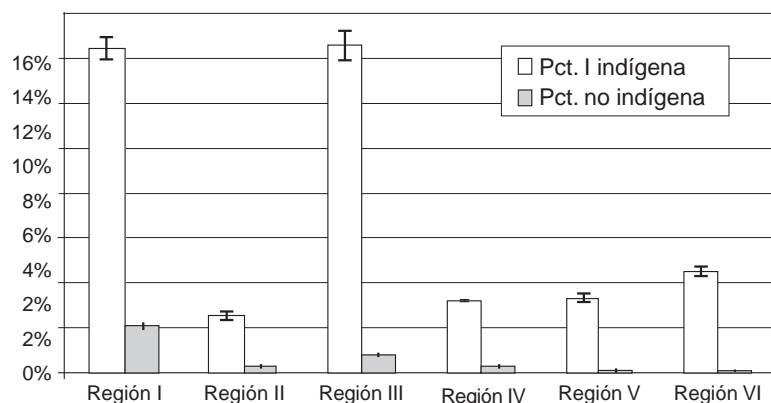
**TABLA 10:**  
**Proporción indígena de víctimas en la Tabla 9a**  
**con etnia conocida**

	Región I	Región II	Región III	Región IV	Región V	Región VI
Proporción indígena	97.8%	99.6%	98.8%	98.5%	99.3%	99.4%
Proporción de víctimas con etnia conocida	24.7%	64.7%	39.4%	99.1%	57.7%	48.4%

G. Con la información de Tablas 8, 9 y 10, se saca la proporción de cada etnia con su error estándar para cada región. Los datos, presentados en la Gráfica 3, explican *inter alia* que según los datos documentados por la CEH, el CIIDH, y el REMHI, más del 14% de la población indígena del área Ixil, de 1981, fue asesinado entre 1981-1983, mientras en el mismo período y área, fue asesinado el 2% de la población no-indígena.

H. Nótese que los datos presentados para la Región VI (área Zacualpa) en la Gráfica 3 no corresponde exactamente a la cifra presentada en el apartado de genocidio porque la definición de la región IV empleada aquí incluye los municipios de Chiche, Joyabaj y Zacualpa. En el apartado de genocidio se incluye únicamente el municipio de Zacualpa. Por lo que la cifras en Gráfica 3 y el apartado de genocidio para región VI fueron calculados sobre bases diferentes aunque con la misma metodología.

**GRÁFICA 3:**  
**Proporción de la población víctima de asesinato documentado, cometidos por fuerzas estatales en Guatemala 1981-1983, por región y grupo étnico, con rango de confianza de 95%.<sup>19</sup>**



I. Los totales proyectados ( $\hat{n}$ ) de cada etnia en cada región fueron calculados por las mismas metodologías descritas en las ecuaciones 2-6, y con los mismos datos que son la base de la Tabla 10. Las cifras se presentan en la Tabla 11, y resumidas en forma *per capita* en la Gráfica 4.

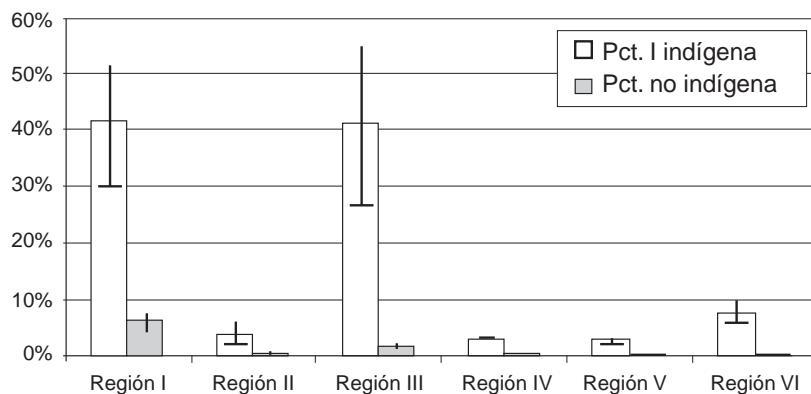
**TABLA 11A: Número de asesinatos proyectados ( $\hat{n}$ ) en las tres bases de datos, por grupo étnico, 1981-1983**

		Región I	Región II	Región III	Región IV	Región V	Región VI
Indígena	$\hat{n}$	2578	443	2983	1012	1020	2723
	$SE(\hat{n})$	190.28	39.13	587.58	0.67	17.11	386.43
No-Indígena	$\hat{n}$	63	2	13	15.7	7.5	6
	$SE(\hat{n})$	5.4	0.07	0.33	0.14	0.12	0.13
Sin etnia conocida	$\hat{n}$	14014.6	394	4791	10	752	1208
	$SE(\hat{n})$	1841.44	151.12	874.1	0.64	31.95	46.79

**TABLA 11B: Número de asesinatos proyectados ( $\hat{n}$ ) en las tres bases de datos, por grupo étnico incluyendo víctimas sin identificar, 1981-1983**

		Región I	Región II	Región III	Región IV	Región V	Región VI
Indígena	$\hat{n}$	16,284	835	7,717	1,022	1,767	3,924
	$SE(\hat{n})$	1,811.0	155.5	1,044.5	0.9	36.1	389.2
No-Indígena	$\hat{n}$	371	4	70	16	13	13
	$SE(\hat{n})$	40.9	0.6	10.5	0.1	0.3	0.3

**GRÁFICA 4:**  
**Proporción proyectada de la población víctima**  
**de asesinatos cometidos por fuerzas estatales en Guatemala 1981-1983,**  
**por región y grupo étnico, con rango de confianza de 95%**



**45.** Nótese que las gráficas 3 y 4 tienen dos interpretaciones fundamentales.<sup>20</sup> Primera: las regiones I y III fueron las más afectadas por la violencia estatal. En los dos regiones hay señales claras de que, en términos cuantitativos, la matanza fue tan masiva que llega a ser genocidio. Segunda: en todas las regiones las víctimas son, desproporcionalmente, personas indígenas. En las barras por región de la Gráfica 4, en la I, se estima que más del 40% de la población indígena fue asesinada mientras 8% de la población no-indígena sufrió lo mismo; la diferencia entre las dos etnias representa un factor de 5, lo que indica que algo, en la estructura de la violencia cometida por el Estado guatemalteco, variaba cualitativamente entre las dos etnias.

## V. Comparaciones entre las bases de datos

**46.** En el análisis de bases de datos múltiples, se puede comparar las bases para determinar los niveles de acuerdo intermuestra en términos temporales, geográficos, y de responsabilidad atribuida a los hechores. Esta parte profundiza la consideración de tendencias y cifras en común entre los tres proyectos.

**47.** En la Tabla 1, es claro que el nivel de acuerdo geográfico es intermedio. En unas regiones los tres proyectos encontraron muchas violaciones, como en regiones 0, I, y III, mientras en otras regiones solamente dos proyectos las cubrieron a fondo (por ejemplo, regiones V, VI, IX), mientras en la región IV únicamente la CEH realizó una investigación intensa.

**48.** Aunque los proyectos cubrieron áreas distintas, el segundo apartado

de esta parte muestra que las tres bases de datos todavía coinciden en la identificación de los meses en que ocurrieron los picos de violencia en el país. El tercer apartado profundiza el análisis de los resultados de los tres proyectos sobre las cifras de responsabilidad para los hechos.

### *Coincidencia en tiempo*

**49.** Si se ordena los datos mensuales del total de asesinatos en cada una de las tres bases de datos, se encuentra un nivel de acuerdo muy alto. En la Tabla 12 los meses fueron ordenados por cantidad de asesinatos contenidos en cada base, presentándose con los porcentajes del total de asesinatos, de los diez meses más violentos en el período 1979-1984.

**TABLA 12:**  
**Los diez meses más violentos**  
**en tres bases de datos, 1979-1984<sup>21</sup>**

Rango	CEH			CIIDH			REMHI		
	Mes	Total	Pctje.	Mes	Total	Pctje.	Mes	Total	Pctje.
1	82-01	2,256	9%	82-02	610	12%	82-03	1,330	12%
2	82-03	2,253	9%	81-06	390	7%	82-02	807	7%
3	82-02	1,880	8%	83-03	297	6%	82-07	792	7%
4	82-08	1,819	8%	82-06	279	5%	82-05	657	6%
5	82-07	1,719	7%	82-07	234	4%	81-09	629	6%
6	81-01	1,423	6%	82-01	233	4%	82-01	470	4%
7	82-06	1,146	5%	82-04	222	4%	82-04	428	4%
8	82-04	937	4%	82-05	210	4%	81-07	397	4%
9	82-05	895	4%	83-08	180	3%	80-02	364	3%
10	81-09	754	3%	81-02	174	3%	82-10	360	3%
Total de 10 meses		15,082	63%		2,829	54%		6,234	56%
Total de 1979-1984		23,890	100%		5,275	100%		11,065	100%

**50.** Los cinco meses en sombra y negrilla son los que coinciden en las tres bases, es decir que entre los peores diez meses, están de acuerdo en cinco meses. Cabe mencionar que incluyen enero, febrero, abril, y mayo del año 1982; otros meses del mencionado año (marzo y junio) coinciden en dos de las tres bases.

**51.** En los peores diez meses de la guerra, las bases coinciden en que sucedieron un poco más de la mitad de los asesinatos documentados (63%, 54% y 56%). La concentración de más de la mitad de las violaciones en esos diez meses está en acuerdo con la Ley de Pareto (en que 80% del fenómeno pasará en 20% de las categorías). Lo que llama la atención es la

cercanía de los meses en tiempo: que los meses que caen en todas las bases ocurrieron en el primer semestre de 1982. Esta concentración es una fuerte evidencia de que lo más intenso de la violencia política en Guatemala sucedió en el mencionado momento. También el nivel de acuerdo entre las bases implica que aunque no investigaron en exactamente las mismas regiones, han encontrado la misma tendencia temporal.

### *Coincidencia en la atribución de responsabilidad*

**52.** Las tres fuentes están fundamentalmente de acuerdo sobre las tasas de responsabilidad atribuidas a las Partes: las tres atribuyen más de 94% de las violaciones con hechos al Estado, y menos de 6% a la guerrilla.

**TABLA 13:**  
**Total de asesinatos con hechor identificado,**  
**por la entidad responsable<sup>22</sup>**

	CEH	CIIDH	REMHI
Estado	24121	2916	19177
Guerrilla	1263	61	1184

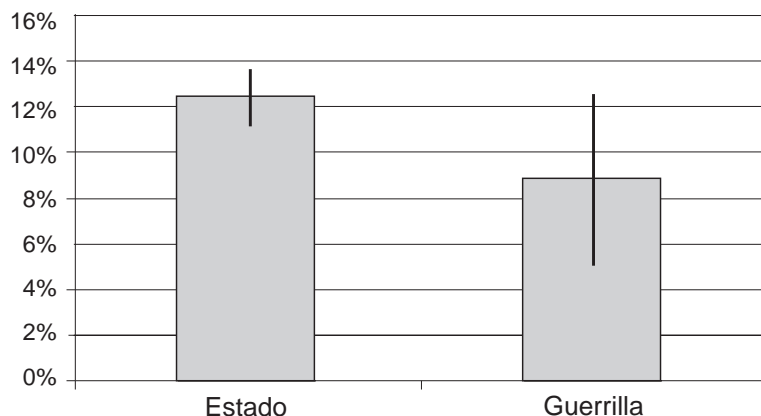
**53.** En el análisis de proporciones de responsabilidad derivadas de muestras que se tomaron no probabilísticamente, como se muestra en Tabla 13 arriba, existe la posibilidad que los proyectos se enfocaran más en los hechos de una de las partes que a la otra. Este especie de “sobre-enfoque” se puede sesgar las proporciones.

**54.** Por ejemplo, si los proyectos se dedicaron a buscar violaciones cometidas por la guerrilla, mientras empleaban menos recursos a la búsqueda para violaciones atribuidas al Estado, este proceso produciría un nivel inflado de responsabilidad atribuida a la guerrilla. En una muestra aleatoria, se evitaría este sesgo al aceptar testimonios seleccionados al azar. Ninguno de los tres proyectos buscó testimonios con un diseño probabilístico en la mencionada forma. Así que hay una cuestión de desproporcionalidad de enfoque en los proyectos que debe ser considerada.

**55.** Los tres proyectos y las medidas de “traslape” explicadas en las Partes II y III, se puede probar la hipótesis que existe sobre-enfoque en uno de los proyectos. Se toman todas las estimaciones de los componentes del total sin duplicación ( $N_k$ ), y se distingue entre los asesinatos atribuidos a la guerrilla y los atribuidos a otras instancias (casi todos, al Estado). Se suman los componentes que “indican” traslape ( $N_{111}+N_{110}+N_{101}+N_{011}$ ) y se dividen

por  $N_k$ ; esta cifra da el porcentaje de traslape por grupo de responsables. Los resultados de este cálculo estan abajo en Gráfica 5.<sup>23</sup>

**GRÁFICA 5:**  
**Tasa de “traslape” de víctimas de asesinato documentado por la CEH, el CIIDH, y el REMHI, por violaciones cometidas por el Estado y por la guerrilla (con barra del rango de confianza 95%)**



**56.** Con los resultados de la Gráfica 4, se puede descartar la hipótesis de que existía una gran diferencia en el nivel de cobertura entre violaciones cometidas por la guerrilla y las cometidas por otras instancias. Aunque hubiera una diferencia pequeña entre el nivel de “traslape” de violaciones sin la presencia de la guerrilla (12.4%) y el de violaciones con la presencia de la guerrilla (8.8%), el error estándar cierre las diferencias a un nivel no interpretable.<sup>24</sup>

**57.** La implicación de la Gráfica 4 es que las violaciones de la guerrilla y las violaciones cometidas por las instancias estatales fueron investigadas con aproximadamente el mismo nivel de cobertura o intensidad, y no existía una desproporcionalidad sistemática en intensidad de investigación sobre las partes suficiente como para cambiar la interpretación de proporciones cuantitativas de responsabilidad.<sup>25</sup>

## *Anexo*

### *Definiciones de regiones, por municipio*

Las áreas fuera de estas definiciones están clasificados como región 0. Note que la definición de región VI en este estudio **no es igual** a la defini-

ción utilizada en el informe de la CEH: este estudio incluyó dos municipios que no se incluyen en el apartado sobre genocidio.

<b>Región</b>	<b>Depto</b>	<b>Municipio</b>
Región I	Quiché	Chajul
Región I	Quiché	San Juan Cotzal
Región I	Quiché	Nebaj
Región II	Alta Verapaz	Cahabón
Región III	Baja Verapaz	Rabinal
Región IV	Chimaltenango	San Martín Jilotepeque
Región V	Huehuetenango	Nenton
Región V	Huehuetenango	San Mateo Ixtatán
Región V	Huehuetenango	Barillas
Región VI	Quiché	Chiche
Región VI	Quiché	Zacualpa
Región VI	Quiché	Joyabaj
Región VII	Guatemala	Guatemala
Región VII	Guatemala	Mixco
Región VIII	Alta Verapaz	Panzós
Región VIII	Alta Verapaz	San Pedro Carchá
Región IX	Quiché	Ixcán
Región X	Santa Rosa	Cuilapa
Región X	Santa Rosa	Barberena
Región X	Santa Rosa	Casillas
Región X	Santa Rosa	Santa Rosa De Lima
Región X	Santa Rosa	Oratorio
Región X	Santa Rosa	San Rafael Las Flores
Región X	Santa Rosa	Santa María Ixhuatan
Región X	Santa Rosa	Taxisco
Región X	Santa Rosa	Chiquimulilla
Región X	Santa Rosa	San Juan Tecuaco
Región X	Santa Rosa	Guazacapán
Región X	Santa Rosa	Naranjo
Región X	Santa Rosa	Pueblo Nuevo Las Viñas
Región X	Santa Rosa	Nueva Santa Rosa
Región X	Escuintla	Escuintla
Región X	Escuintla	Santa Lucía Cotzumalguapa
Región X	Escuintla	La Democracia

Región X	Escuintla	Siquinalá
Región X	Escuintla	Masagua
Región X	Escuintla	Tiquisate
Región X	Escuintla	La Gomera
Región X	Escuintla	Guanagazapa
Región X	Escuintla	San José
Región X	Escuintla	Iztapa
Región X	Escuintla	Palín
Región X	Escuintla	San Vicente Pacaya
Región X	Escuintla	Nueva Concepción
Región X	Retalhuleu	Retalhuleu
Región X	Retalhuleu	San Sebastián
Región X	Retalhuleu	Santa Cruz Mulua
Región X	Retalhuleu	San Martín Zapotitlán
Región X	Retalhuleu	San Felipe
Región X	Retalhuleu	San Andrés Villa Seca
Región X	Retalhuleu	Champerico
Región X	Retalhuleu	Nuevo San Carlos
Región X	Retalhuleu	El Asintal
Región X	San Marcos	Nuevo Progreso
Región X	San Marcos	El Tumbador
Región X	San Marcos	Malacatán
Región X	San Marcos	Catarina
Región X	San Marcos	Ayutla
Región X	San Marcos	Ocos
Región X	San Marcos	Pajapita

---

**NOTAS:**

1. A solicitud de la CEH, este análisis fue conducido por el Dr. Patrick Ball, Subdirector del Programa en Ciencia y Derechos Humanos de la Asociación Norteamericana para el Avance de la Ciencia (AAAS), con la asesoría de Dr. Herbert Spírer ("Adjunct Professor", Columbia University y Consultor de la AAAS), Dr. Frederick Scheuren ("Senior Fellow", The Urban Institute, y "Adjunct Professor", George Washington University), y William Seltzer ("Senior Researcher", Fordham University).
2. Se excluyen de la Tabla 1 las víctimas cuando la fecha de violación no estuviese identificada con precisión hasta el año y aquellas sin un lugar de la violación identificada con precisión hasta el municipio.
3. Aunque el CIIDH también recogió información de fuentes periodísticas y de fuentes documentales, en este análisis únicamente se incluyó la información derivada de fuentes directas, y con la firma del declarante.
4. Cabe mencionar que la definición de Región VI (área Zacualpa) incluye los municipios de Chiché y Joyabaj, y no corresponde exactamente con con la definición de la región Zacualpa en el apartado del informe que trata genocidio que incluye solamente el municipio de Zacualpa.

5. Las Regiones fueron definidas para aislar áreas en que existían grandes diferencias en el nivel de cobertura entre los proyectos.
6. Note que no se trata de desapariciones forzadas (las que son contabilizadas aparte). No hubo tiempo ni recursos para extender este análisis a desaparecidos.
7. Nótese que este análisis trata víctimas, no violaciones, pero para la violación de asesinato, las dos medidas son iguales porque cada víctima puede sufrir asesinato solamente una vez. Véase a *Who Did What to Whom? Designing and Implementing a Large-Scale Human Rights Data Project*, Patrick Ball (Washington: AAAS), 1996.
8. Además, muchas víctimas no están identificadas por nombre en las bases de datos. Los testigos originales mencionaron un grupo de personas. Testigos distintos con frecuencia estiman números desiguales en hechos que involucran a muchas personas, no es posible concordar víctimas no identificadas. Este análisis presume que las tasas de concordancia entre las víctimas no identificadas son iguales a las tasas entre las identificadas.
9. Las muestras fueron estratificadas proporcionalmente por región para asegurar la cobertura a todas.
10. De los registros mencionados en el texto, unos 498 registros más fueron seleccionados y concordados una segunda vez. Nos referimos a estos registros en el análisis de la calidad de la concordancia.
11. En el análisis de concordancia, lo que nos preocupa es que los registros que realmente son concordados se escapen a los analistas. Los 498 registros concordados dos veces no fueron al azar, sino fueron seleccionados para representar registros ya concordados para dar una prueba de la eficacia y consistencia del proceso. Comparando los 171 registros, tratados por los dos analistas, que fueron clasificados por uno o los dos como un registro concordado, 88% fueron codificados iguales.
12. Esta explicación está tomada de *Population Growth Estimation: a Handbook of Vital Statistics Measurement*, por Eli S. Marks, William Seltzer, y Karol J. Krótki (New York: The Population Council), 1974, pg. 13-17.
13. Véase ecuación 7.188 Marks, Seltzer y Krótki, pg. 406.
14. Tomado de Kirk M. Wolter, *Introduction to Variance Estimation* (New York: Springer-Verlag), 1985, pg. 154-155.  $\hat{\theta}_r$
15. Los "pseudovalores"  $\hat{\theta}_r$  deben ser aproximadamente independientes y distribuidos idénticamente. Esta presunción fue probada con un gráfico de probabilidad normal para cada vector de pseudovalores, y en cada caso fueron bien distribuidos.
16. La estimación de  $N_i$  que fue de 47,706 asesinatos documentados en el proyecto, junto con su error estándar 228, tiene un rango de confianza de 95% 47,559 – 48,152. Note que este rango de confianza incluye el valor estimado en la Tabla 5, 47,803. La cercanía del valor de la Tabla 5 con el valor estimado por la suma de las Regiones por el método "jackknife" implica que no hubiese tanto sesgo en la estimación simple. No obstante, el sesgo preocupante que requirió la desagregación a Regiones y el uso del método de "jackknife" no afectaba a  $N_i$ , sino que afectaría mucho a  $N_{1000}$ , y entonces fue todavía necesario.
17. Veáanse, en ese contexto, el Informe del Proyecto REMHI, *Guatemala: Nunca Más Tomo IV, Víctimas del Conflicto* (Guatemala: Oficina de Derechos Humanos del Arzobispado de Guatemala) 1998, pg. XXXI-XXXII; y Patrick Ball, Paul Kobrak, y Herbert F. Spierer, *Violencia Institucional en Guatemala, 1960 a 1996: una Reflexión Cuantitativa* (Washington DC: AAAS), 1999, pg. 62, nota 12.
18. Las tasas de traslape no fueron calculadas por grupo étnico. En cambio, las tasas Regionales para los concordados del período 1981-1983 fueron aplicados a los tres grupos étnicos en cada Región. Esta aplicación presume que las tasas de "traslape" no varían (en términos significativos) entre grupos étnicos.
19. Fuente de la gráfica: censo de 1981; testimonios recibidos por la CEH, fuentes directas al CIIDH, y testimonios recibidos por el REMHI.
20. También cabe mencionar que, en términos absolutos y relativos, el error estándar de cada cifra en la Gráfica 3 es mucho más que el error estándar en Gráfica 2. Esta diferencia es correcta, dada que las proyecciones en la Gráfica 3 incorporan más incertidumbre que las estimaciones en la Gráfica 2. Los tamaños de las muestras en que están basadas las estimaciones de los totales sin duplicación ( $N_i$ ) son adecuados para estimaciones de  $N_i$  sin mucho error. La proyección todavía tiene más incertidumbre que la estimación de  $N_i$ , reflejada en el error más alto.
21. En la Tabla 12, solamente se incluyen los asesinatos que son identificados por fecha con precisión hasta el mes.
22. Se incluye solamente aquellas violaciones con responsable identificado y con la fecha de la violación con precisión hasta el año. Otra vez cabe subrayar que este estudio contabiliza asesinatos, no solamente ejecuciones extrajudiciales.

23. La tabla en que está basada la Gráfica 5 (corresponde a Tabla 5, dividida por entidad responsable) se muestra a continuación.

Categoría	Estado	Guerrilla
N <sub>111</sub>	299	3
N <sub>110</sub>	617	
N <sub>101</sub>	3769	200
N <sub>011</sub>	388	
N <sub>100</sub>	19,173	1,087
N <sub>010</sub>	2,165	61
N <sub>001</sub>	14,430	949
N <sub>k</sub>	40,842	2,301
Tasa de traslape $\frac{N_{111} + N_{110} + N_{101} + N_{011}}{N_k}$	12.4%	8.8%
SE (N <sub>k</sub> )	0.5%	1.9%

El error estandar está calculado por el método convencional para proporciones derivadas por muestras,

$$SE = \sqrt{\frac{P * (1 - P)}{N}}, \text{ y el rango de confianza } +/- 1.96 * SE.$$

24. La diferencia entre la tasa de "traslape" para asesinatos cometidos por el Estado y por la guerrilla no es significativo, ni en términos puramente técnicos, no es significativo en términos analíticos. Para el análisis del efecto en las proporciones estimadas, veáse la nota 25, abajo. La prueba técnica es la siguiente:

$$SE = \sqrt{\frac{P_E * (1 - P_E)}{N_E} + \frac{P_G * (1 - P_G)}{N_G}}, = 0.0193, \text{ que se da un rango de confianza } +/- 3.8\%. \text{ La diferen-}$$

cia entre las dos tasas es 12.4% - 8.8% = 3.6%; el rango de confianza es más que la diferencia, que implica que no se puede descartar la hipótesis que la diferencia es igual a cero. Este cálculo confirma lo que es claro en términos intuitivos en Gráfica 5.

25. Hay dos maneras de considerar el efecto de las tasas de traslape en la proporción de responsabilidad atribuidas al Estado y a la guerrilla. La proporciones de responsabilidad atribuidas a las dos partes como resultado de los N<sub>k</sub> estimados en nota 23, arriba, están presentadas abajo; nótese que son un promedio de las proporciones de las tres bases de datos, como las implicadas en Tabla 13 en el texto.

Estimación	Estado	Guerrilla
N <sub>k</sub>	41,147	1,860
Proporción del total N <sub>k</sub>	95.7%	4.3%
n <sub>1000</sub>	73,622	3,706
$\hat{N}$	114,769	5,567
Proporción del total $\hat{N}$	95.4%	4.6%

Para ver el efecto de la pequeña desproporcionalidad en cobertura, se calculó n<sub>1000</sub> (se usa la Ecuación 3 para el Estado, pero se usa la Ecuación 2 para la guerrilla porque hubiera datos de solamente dos proyectos). La estimación de n<sub>1000</sub> incluye la información sobre tasas de "traslape", y por eso n<sub>1000</sub> controla el efecto de la desproporcionalidad en cobertura. Nótese que las proporciones calculadas por  $\hat{N}$  resultan iguales a las calculadas con N<sub>k</sub>. La conclusión es que la desproporcionalidad en cobertura sobre los hechos de las Partes no cambiaron el análisis final de su responsabilidad.

*Guatemala, memoria del silencio* terminó de imprimirse en el mes de junio de 1999, en los talleres de Litoprint, 3a. Calle "A" 37-47 Zona 7, Guatemala 01007.

*Guatemala, memoria del silencio* terminó de imprimirse en el mes de junio de 1999, en los talleres de Litoprint, 3a. Calle "A" 37-47 Zona 7, Guatemala 01007.

*Guatemala, memoria del silencio* terminó de imprimirse en el mes de junio de 1999, en los talleres de Litoprint, 3a. Calle "A" 37-47 Zona 7, Guatemala 01007.

*Guatemala, memoria del silencio* terminó de imprimirse en el mes de junio de 1999, en los talleres de Litoprint, 3a. Calle "A" 37-47 Zona 7, Guatemala 01007.