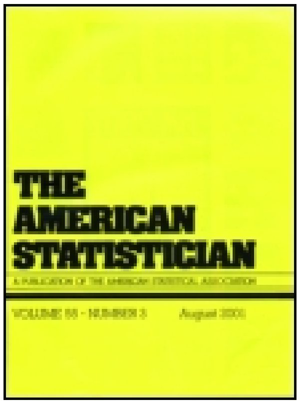

The American Statistician
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/utas20

Applications of Multiple Systems Estimation in Human Rights Research

Kristian Lum [a], Megan Emily Price [b] & David Banks [c]

[a] Network Dynamics and Simulation Science Laboratory, Virginia Tech, Blacksburg, VA, 24061

[b] Benetech, Palo Alto, CA, 94306

[c] Department of Statistical Science, Duke University, Durham, NC, 27708

Published online: 18 Nov 2013.

Applications of Multiple Systems Estimation in Human Rights Research

Kristian LUM, Megan Emily PRICE, and David BANKS

Multiple systems estimation (MSE) is becoming an increasingly common approach for exploratory study of underreported events in the field of quantitative human rights. In this context, it is used to estimate the number of people who died as a result of political unrest when it is believed that many of those who died or disappeared were never reported. MSE relies upon several assumptions, each of which may be slightly or significantly violated in particular applications. This article outlines the evolution of the application of MSE to human rights research through the use of three case studies: Guatemala, Peru, and Colombia. Each of these cases presents distinct challenges to the MSE method. Motivated by these applications, we describe new methodology for assessing the impact of violated assumptions in MSE. Our approach uses simulations to explore the cumulative magnitude of errors introduced by violation of the model assumptions at each stage in the analysis.

KEY WORDS: Capture-recapture; Log-linear models; Mortality estimates; Record linkage.

1. INTRODUCTION

Recently, there have been new efforts to prosecute crimes against humanity in national and international courts (Ball et al. 2002; Guzmán 2011); examples include Charles Taylor of Liberia, Augusto Pinochet of Chile, Laurent Gbagbo of the Ivory Coast, and Ríoss Montt of Guatemala. Truth and Reconciliation Commissions have emerged as a social tool for coping with past injustices, and these depend upon defensible estimates of magnitude and culpability (Ball et al. 2003). These trends have forced human rights workers to adopt more rigorous methodology in counting the dead, the disappeared, and the damaged. When naïve estimates are unreliable, public discussion can be hijacked for political ends (see, e.g., the controversy over the number of deaths in the Darfur conflict, at *http://news.bbc.co.uk/2/hi/africa/6951672.stm*).

Besides these prominent applications, governments, political scientists, and historians also need to study trends and patterns in violence. Governments often use estimates of human rights abuse as a component in decision-making about foreign aid and development policies. Researchers want to understand the evolution of ethnic or religious hatred and how this interacts with geography, social class, and gender. Such studies demand defensible estimates of conflict mortality and related social costs.

To these ends, statisticians have powerful tools. In this article, we focus on multiple systems estimation (MSE) and describe several cases in which this particular class of methods played an essential role in assessing large-scale human rights violations. Broadly speaking, MSE methods seek to estimate the total size of a population based on multiple data sources and the overlaps among them. MSE was initially developed as capture-recapture methodology in ecology (Peterson 1896; Lincoln 1930). During the ensuing century, MSE methods have been applied to human research subjects in demography (Sekar and Deming 1949; Seber 1965; Marks, Seltzer, and Krótki 1974; Darroch et al. 1993), epidemiology (Wittes and Sidel 1968; Wittes, Colton, and Sidel 1974; International Working Group for Disease Monitoring and Forecasting 1995a,b; Hook and Regal 2000; Seber, Huakau, and Simmons 2000), and, in just the past two decades, human rights (Ball 2000; Ball et al. 2002, 2003; Brunborg, Lynstad, and Urdal 2003; Silva and Ball 2006; Lum et al. 2010; Zwierzchowski and Tabeau 2010). This class of methods has been modified and expanded as new fields of applications pose unique challenges to the underlying methodological assumptions. In the case of human rights research, MSE allows statisticians to make compelling inferences about human rights abuses that could not be obtained from simple tabulations.

Humanity wants to avert genocides, arrest violent outbreaks, and rebuild societies that have been divided by conflict. MSE, in conjunction with emerging work in economics (Far 2005), sociology (Turner 2006), public health (Cairns et al. 2009; Roberts et al. 2010), and political science (Davenport and Ball 2002), can help understand the social processes that lead to conflict. The three case studies presented in Section 3 span over a decade of human rights research and show the increasing sophistication of MSE methods. The Guatemala case, with data from 1978 to 1996, is the earliest application and its analysis used very

Kristian Lum is Assistant Research Professor and Statistician, Network Dynamics and Simulation Science Laboratory, Virginia Tech, Blacksburg, VA 24061 (E-mail: *klum@vbi.vt.edu*). Megan Emily Price, Benetech, Palo Alto, CA 94306 (E-mail: *meganp@benetech.org*). David Banks is Professor, Department of Statistical Science, Duke University, Durham, NC 27708 (E-mail: *banks@stat.duke.edu*). The authors thank Patrick Ball and Jeff Klingner for their helpful suggestions. This work has been partially funded by Virginia Tech internal funds.

different techniques than our most recent study, based on data collected in Colombia through 2008. These case studies are used to show how progress in MSE theory has enabled analysts to relax their reliance on potentially problematic modeling assumptions. Section 2 provides an overview of MSE and traces its historical development, highlighting new work that has improved MSE's robustness and accuracy while pointing out issues that remain unsolved; Section 3 summarizes three accounts of the use of MSE in the field of human rights. Section 4 presents a small simulation study that assesses the impact of differing list coverage on MSE estimates. Finally, Section 5 closes with conclusions based on the three case studies and small simulation example.

2. MULTIPLE SYSTEMS ESTIMATION

MSE is a class of techniques that seeks to estimate the total size of a population using multiple data sources and the overlaps among them. In the field of human rights, the goal of MSE is to understand patterns of violence that are not fully apparent in the *reported* data. In most human rights applications to date (and in all of the case studies presented here), the violations under study are lethal; that is, we are estimating the total number of killings. There are strong reasons to believe that killings occurred that are not included in the reported data, and that these missing data records occurred in such a way that conclusions based on the observed data would be biased (Davenport and Ball 2002). To address the missing data issue, researchers rely upon a method called MSE (often referred to as capture-recapture in other literature) to make inferences about the extent and kind of underregistration of the violations that were committed. As introduced by Peterson (1896), the method was originally developed to estimate a fish population. Despite this early application, the strategy is relevant to human rights applications as well and enables quantification of the uncertainty regarding the number of people whose deaths were not reported.

There are many reasons why a lethal violation might not be reported. As noted by Ball et al. (2003), it may happen that victims live in remote regions where recording agencies do not exist. Some victims or the families of the victims may fear retaliation if the violation is reported. Some violations may occur without any witnesses. And some entail other kinds of violence, such as rape, that the witness may want to conceal. For all these reasons and more, the recorded conflict mortalities represent only a lower bound on the true count. (Strictly speaking, the raw data provides a lower bound only if all records have been successfully matched within and between databases.)

2.1 Data and Notation

In a MSE setting, we typically have d administrative lists from which to estimate the number of names that do not appear on any list. The data from the lists, after being deduplicated and matched, is then collapsed into list intersection counts, $\mathbf{Y} = \{Y_{\mathbf{s}} : s_i \in \{0, 1\}, i = 1, \ldots, d\}$, where $\mathbf{s}$ is a set of d binaries that indicates the intersection. This is, perhaps, best explained by example. Let Y_{111} be the number of names that appear on all three lists in a three-list system; Y_{1001} be the number of names that appear on Lists 1 and 4 only in a four-list system, and so forth. Then $Y_{00\ldots0}$ is precisely what we seek to estimate: the number of names that appear on no list. To denote marginalization, Y_{1+} is the total number of names that appear on List 1, regardless of whether they appear on List 2. Y_{+1} is similarly defined as the total number of names that appear on List 2.

2.2 A Simple Example

As a simple example of two-system estimation, consider Peterson's original application. An initial netting of Y_{1+} fish are caught, tagged, and released. Then, in a second netting, Y_{+1} fish are captured, of which Y_{11} are seen to be tagged (i.e., they were recaptured). Under the assumptions that

1. there are no births, deaths, emigrations, or immigrations (i.e., the pond is a closed population),
2. within each netting, all fish are equally likely to be caught (homogeneous capture probability),
3. being captured on the first netting does not affect the probability of being captured on the second; in particular, fish that are caught do not become more cautious (independent systems), and
4. the tagging is error free (perfect matching),

it is simple to obtain the point estimate of the number of fish in the pond as $Y_{1+}Y_{+1}/Y_{11}$, and also the corresponding variance of that estimate.

Obviously, people are not fish, and some of these assumptions can fail significantly in the human rights applications we consider. The exception is the closed population assumption, because the dead stay dead. But each of the other assumptions can be problematic to varying degrees, depending upon the specific situation. In the following sections, we discuss ways in which assumptions (2) and (3) can be relaxed and how assumption (4) is typically handled.

2.3 Heterogeneous Capture Probabilities

If we assume homogeneous capture probabilities, then our model specifies that within each data collection effort, all violations have equal probability of being recorded. Of course, distinct systems may have different probabilities; for example, a well-funded survey will capture more violations than an underfunded one.

This homogeneity assumption is usually wrong. As mentioned by Bishop, Fienberg, and Holland (1975), "social visibility" is often an issue—victims who have large social networks are more likely to be recorded than others. This was particularly problematic in the Guatemala case study of Section 3.1, when prominent activist clergy or their relatives were killed. Worse, the estimated probability of a violation being recorded varied by location; rural killings were more likely to be reported than urban disappearances (Patrick Ball, personal correspondence, 12 April 2011). In the Guatemala analysis, it was possible to stratify by ethnicity and location, and within strata the assumption of homogeneity was more tenable. A side benefit of stratification was finer-grained insight into the role of ethnicity, region, and class in the conflict.

An alternative to stratification is a formal model-based treatment of the heterogeneity of individual capture probabilities. Such models are delicate. Huggins (2001), and others, pointed out that in some models the total number of violations is unidentifiable. Nonetheless, Chao (1987) addressed individual heterogeneity of capture probability by assuming a parametric form for the distribution of the individual capture probabilities. Chao, Lee, and Jeng (1992) relaxed the assumption that all of the lists have the same capture probability for each individual, letting the capture probability for the ith person on the jth list be the product of the ith individual's capture probability and an inclusion probability for the jth list. Other methodological approaches include mixture models that treat the heterogeneity component as a random effect (Coull and Agresti 1999). In the cases of both Peru and Colombia, information was available from recent censuses (1981 for Peru, 2005 for Colombia) that could be used to model individual capture probabilities in terms of age, gender, region, and ethnicity.

A different approach that explicitly models individual capture heterogeneity based on Grade of Membership was introduced by Woodbury, Clive, and Arthur Garson (1978). This was fundamentally extended by Manrique-Vallier and Fienberg (2008); their work was partially motivated by field experience in the Peruvian case study. In Grade of Membership analyses, an individual's capture probability on each list is modeled as a mixture of latent class variables. Basu and Ebrahimi (2001) also employed a mixture model that allows for two different classes with distinct capture probability for each class by each system. However, there is a large gap between theory and practice in human rights research.

2.4 System Dependence

System dependence among administrative lists occurs if the appearance of a reported death on one list implies that the same person is more (or less) likely to be reported on another list. This was a major problem in the Colombia case study highlighted in Section 3.3, where Benetech analyzed lists collected by both government agencies and nongovernmental organizations (NGOs). Negative dependence occurred because people who were inclined to report a violation to an NGO were less likely to trust the government (and vice versa). Positive dependence occurred because records of homicides kept by the National Police were partially subsumed in the list of violations collected by the Office of the Vice Presidency, and both lists were among those used in the analysis.

If not adequately accounted for, positive list dependence leads to underestimation of the number of violations, whereas negative dependence results in overestimation. Field workers can often assess certain dependencies from their experience with the various data collection agencies, but that contextual knowledge can typically only address pairwise dependence. Judging higher-order dependence structure among systems is more difficult.

In Peterson's pond example, the assumption that the two nettings were independent was essential. But when there are more than two collections, there is greater flexibility (Fienberg 1972; Cormack 1989). A common log-linear model for MSE is given by

$$E[Y_{\mathbf{s}}] = \ln(m_{\mathbf{s}}) = \alpha + \sum_{i=1}^{d} \beta_i s_i + \sum_{i=1}^{d} \sum_{j>i} \beta_{ij} s_i s_j, \qquad (1)$$

where $\mathbf{s}$ is the set of binaries that indicates list intersection as defined in Section 2.1. For example, if $\mathbf{s} = \{1, 0, 1\}$, then $E[Y_{101}] = \alpha + \beta_1 + \beta_3 + \beta_{13}$. Log-linear models allow for the detection and estimation of list dependencies through the inclusion and estimation of parameters for interaction terms for various sets of lists. In Equation (1) above, the terms in the second summand represent two-way list interactions. Omitting all interaction terms reduces this model to the case of system independence.

These models can be extended to examine heterogeneous capture probabilities, described in the previous section, as shown by Darroch et al. (1993), building on the Rasch model. Fienberg, Johnson, and Junker (1999) expanded on this approach with both classical multilevel and fully Bayesian hierarchical models.

2.5 Model Selection and Averaging

This expanded flexibility in modeling now raises the question of which model to choose. In the log-linear setting, Bishop, Fienberg, and Holland (1975) discussed conditional likelihood tests for determining which log-linear model best describes the dependence structure exhibited by the data. The Peruvian report (Ball et al. 2003) uses a χ^2 test to choose among several three-system models. Alternative model-fit statistics, such as the AIC or BIC, could also be used.

In some cases, such as the Colombia example of Section 3.3, the fit statistics are almost indistinguishably similar for models that produce quite different estimates of N, the total number of individuals in the population. When that happens, it is not sensible to ignore models that are nearly as good as the model found to be "best" by fit statistics, especially if those second-best models produce very different estimates. One solution is to employ Bayesian model averaging (Raftery 1995; King and Brooks 2008). York and Madigan (1992) showed how to use model averaging, as opposed to model selection, in the setting of graphical models to average across list dependence structures based upon the marginal likelihood of each model. Employing this approach generally requires one to enumerate all possible models, and in the Colombia case study, the 15 datasets generate so many models that the problem is computationally intractable.

If one does not restrict the set of models under consideration, aside from requiring that the model contain an intercept and not be overdetermined, then the number of possible models is $2^{2^d - 1} - 2$. For the case in Colombia, in which $d = 15$, this number is impractically large and not all models can be considered. Lum et al. (2010) introduced a method that considers only three systems at a time and averages over all models applied to each three-system partition, with an additional layer that averages over all three-list partitions, to create a tractable model-partition space. Alternatively, Madigan, York, and Allard (1995) explained how to use Markov chain Monte Carlo model composition to average across a universe of models that is not easily enumerable by stochastically walking through the model space, switching from the current model to a new proposed model with probability that is a function of the ratio of the two marginal likelihoods.

2.6 Record-Linkage

The fourth critical assumption for MSE calculations is that matching has been perfect. This is necessary because the size of each list and the size of the overlaps among lists must be known. Therefore, individual records must be both identifiable as unique within a list (e.g., duplicate records must be reconciled and eliminated within a single list) and identifiable when identical across lists (e.g., if the same record appears on multiple lists, it must be recognized as the same).

Matching is particularly challenging in cross-cultural contexts, where names, dates, employment, addresses, and family relationships can have different meanings. For example, in many societies, names are not a distinctive identifier because of common given names or surnames (Silva, Klingner, and Weikart 2010). Frequently, complete victim names are unknown or are intentionally withheld and additional information, such as the date or location of the violation, is necessary to generate confidence in matched records.

Record matching is complex. The initial statistical strategy for record linkage was the model developed by Fellegi and Sunter (1969). What follows is a generalization of that model. Given two records, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, one calculates some (usually vector-valued) statistic $\boldsymbol{h}(\boldsymbol{x}_1, \boldsymbol{x}_2)$ and estimates $\mathrm{P}(\text{match} \mid \boldsymbol{h}(\boldsymbol{x}_1, \boldsymbol{x}_2))$ according to an appropriate model. If this estimated probability exceeds some user-specified value γ, then the algorithm declares that the records are a match.

To make this concrete, suppose there are two records:

$$\boldsymbol{x}_1 = \{\text{R. Munson, 710 Duckpin Lane, Bowling Green, KY}\}$$
$$\boldsymbol{x}_2 = \{\text{Roy Munson, 710 Duckling Street, Bowling Green, KY}\}.$$

The analyst would use their best judgment to create a function $\boldsymbol{h}$ that extracts information from these records (e.g., match on first initial, equal street number, similar street name, same city name, same state) and build a model that describes the probability that these records refer to the same person. That model could be calibrated against, say, a corrupted version of a known database. In principle, depending on the modeling effort, this approach would take account of the fact that "Munson" is a rarer name than "Smith," increasing the estimated match probability, or that Bowling Green is relatively small, also increasing the estimated probability.

Much of the art in this science derives from the selection of the function $\boldsymbol{h}$. Phillips (1990) let $\boldsymbol{h}$ reflect phonetic similarity: Duckpin and Duckling sound a bit alike, so if the information for either record were collected aurally, it could explain the discrepancy. Other strategies look for matches on alphanumeric strings separated by punctuation or focus on distances between short character strings (so that misspellings have little influence); these may involve metrics that are based on "edit distance" or "affine gaps." Elmagarmid, Ipeirotis, and Verykios (2007) provided a recent survey of this literature. In terms of implementation, the WEKA project (Hall et al. 2009) provides machine learning algorithms that are useful for record linkage.

Additional choices exist in the selection of the statistical model for estimating match probabilities. The default analysis uses logistic regression, but more sophisticated record-linkage techniques employ generalized additive models or other strategies. The success of such models can be assessed by application to a real or artificial database in which known amounts and kinds of corruption have been introduced. A good assessment should be able to duplicate the kinds of mistakes that occur in the actual collection process. One advantage of such assessment is that the analyst can tune the simulation to the particulars of their circumstance; record linkage with American names and addresses is a very different problem from linking, based on, say, Colombian names and addresses.

In the context of human rights applications, the challenge is more difficult than linking names and addresses. A report might indicate that a male body was found at a given intersection on a given date; a potential matching record might indicate that a specific male was abducted by armed men a day earlier at a nearby location. Building appropriate summary functions $\boldsymbol{h}$ and matching models is an on-going challenge. Further complications arise if one believes there has been falsification of records, that is, reports of imagined or intentionally invented violations. This is certainly a plausible scenario, particularly in human rights applications, but this issue has gotten little attention to date. One might imagine using the matching process to identify candidate falsified records, as it is unlikely that a falsified record would match records from other agencies without a large-scale conspiracy. This would imply close examination of records in what we refer to as the "singleton" cells—records that appear on one and only one list. The potential impact of such falsified record patterns is examined in the context of one specific example in Section 3.1.

There is no conservative approach to record-linkage. Because MSE is based not only on the total number of uniquely identified names but also on the overlaps among the names on the lists, varying the sensitivity of the matching algorithm does not result in predictable effects on the final MSE estimates. Because of this, it is wise to propagate matching uncertainty through the entire MSE analysis.

For each pair of records, the analysis produces an estimated probability that they match. From this, one can simulate realizations of the match sets among the different systems used to produce the MSE inference. By chance, one simulation, in which records are matched according to their estimated probabilities, might yield a relatively high estimate of duplication; a second simulation might produce a lower estimate. By repeating these simulations many times and producing corresponding point estimates through MSE for each, one can obtain a confidence interval on the number of casualties that takes account of the probabilistic uncertainty in the matching procedure. More sophisticated still, Fienberg and Manrique-Vallier (2009) explored a model that jointly accounts for matching and estimation; Tancredi and Liseo (2011) presented the same in a fully Bayesian framework.

Table 1. Recorded number of killings across datasets in Guatemala

		CEH			
		Yes		No	
		CIIDH			
		Yes	No	Yes	No
REMHI	Yes	393	3943	634	15,955
	No	898	19,663	6,317	Y_{000} (?)

3. CASE STUDIES

In this section, we provide the historical setting of three situations in which MSE was used to estimate the magnitude of conflict-related deaths, the data available for each analysis, and the questions that motivated each project. These cases are not meant to be exhaustive but rather to highlight the role that such analyses can play in human rights advocacy work and to examine the challenges of such methodology in this field. In particular, these applications illustrate cases for which researchers can plausibly anticipate distinct kinds of failures in the model assumptions needed for traditional MSE.

The first case is an analysis of the number of killings that occurred in Guatemala during that country's 36 years of internal armed conflict. This analysis marks the first time that MSE was used to address the extent of undercounting of human rights violations. The second case is from Peru, where MSE estimates tripled the previously generally accepted estimate of killings. In Colombia, the number and diversity of available datasets posed a novel challenge for established MSE methodology. But a modified MSE technique developed for this situation found large numbers of undocumented killings and disappearances. Finally, although we do not showcase it here, the interested reader might also find the use of MSE in Kosovo applicable (Ball et al. 2002).

3.1 Guatemala

Beginning with the attempted overthrow of the government of General Miguel Ydígoras Fuentes by army officers in 1960, an internal armed conflict raged in Guatemala from 1960 to 1996. During this time, the Guatemalan government carried out a massive counter-insurgency campaign of extra-judicial killings and disappearances. The government's main opponents in the conflict were guerilla groups that formed to combat government coercion, workers' rights groups, and university intellectuals, who spoke out against the oppressive policies (Ball, Kobrak, and Spirer 1999).

Three databases describe the violence in Guatemala between 1960 and 1996 based primarily on victim and witness testimonies. These databases are from projects conducted by the Commission for Historical Clarification (CEH), the Recovery of Historical Memory (REMHI), and the International Center for Human Rights Investigations (CIIDH[1]). The CIIDH database contains multiple sources, but only testimonies were used in MSE analyses (Ball 2001).

The CEH asked researchers at the American Association for the Advancement of Science (AAAS) to analyze the three datasets to answer the question "How many people were killed in Guatemala during the period of the CEH mandate, 1960–1996?" (Ball 1999, chap. 11). Estimates were actually calculated for a shorter time period, 1978–1996, due to data sparsity for the earlier years. Table 1 summarizes the number of killings documented in each dataset.

After matching and deduplicating among the three lists (see Section 2.6), Ball (1999, chap. 11) estimated a total of 47,803 *reported* killings. Using the method suggested by Marks, Seltzer, and Krótki (1974) for cases where possible correlation bias is suspected, Ball (1999, chap. 11) estimated that approximately 84,468 killings were not reported to any project, for a total of 132,174 killings in Guatemala between 1978 and 1996, with a standard error of 6568.

More specifically, the unobserved count was estimated by

$$E[Y_{000}] = m_{000} = \frac{Y_{100}Y_{010} + Y_{100}Y_{001} + Y_{010}Y_{001}}{Y_{110} + Y_{101} + Y_{011}}, \qquad (2)$$

where m_{000} is expected number of victims who were not reported to any of the three projects. The standard error of the estimate was calculated using the jackknife method (see Ball 1999, chap. 11, for full details).

Even more important than the estimated magnitude of total killings is the calculation of this estimate by ethnic group and region. The Commission for Historical Clarification (CEH) used secondary sources and qualitative evidence to identify six geographic regions during 1981–1983 in which they believed state violence was concentrated against indigenous people. Using the same MSE method described above, separate estimates were calculated for each ethnic group in each region for the years 1981–1983. The 1981 census was then used to calculate a killing rate. As shown in Figure 1, in at least two of the six regions, indigenous people (darker gray bars) were killed at a significantly higher rate than nonindigenous (lighter gray bars) people. In fact, in these regions it was estimated that 40% of indigenous people had been killed, a rate five to eight times greater than that of the nonindigenous people. The Commission for Historical Clarification used this, among other evidence, as an indication that acts of genocide were

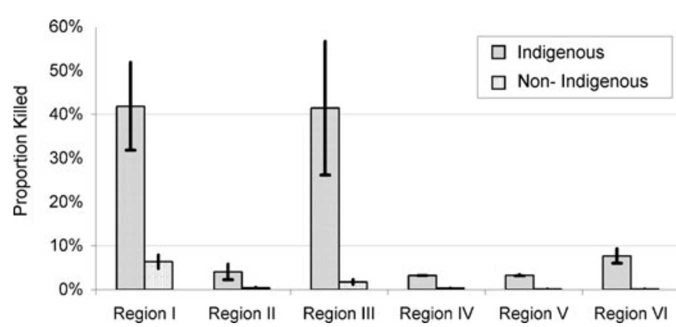

Figure 1. Estimated proportion of ethnic groups killed in Guatemala, 1981–1983 (figure modified from Ball 1999, chap. 11). Black bars indicate confidence intervals.

[1] All acronyms reflect project names in Spanish.

committed against the indigenous people of Guatemala (Ball 2001).

This analysis was recently revisited, motivated by the trial of General José Efraín Ríos Montt and General José Mauricio Rodríguez Sánchez for genocide and crimes against humanity.[2] This updated analysis included a fourth dataset and an examination of the possibility of fabricated records. In this case one particular dataset was considered less reliable than the other three, so analyses were conducted dropping between 5% and 50% of records of indigenous deaths reported *only* to this dataset. In other words, analyses were replicated under scenarios where varying proportions of one list were considered to be potentially incorrect and thus excluded from calculations. This analysis found no change in the substantive conclusion that indigenous people were killed at a much higher rate than nonindigenous people in these regions at these times. Including all records resulted in an estimated indigenous mortality rate nearly eight times the mortality rate of nonindigenous; dropping half of the unique records of indigenous deaths from one dataset resulted in an estimated indigenous mortality rate nearly five times the rate of nonindigenous.[3]

3.2 Peru

Between 1980 and the mid-1990s, the Sendero Luminoso, or "Shining Path," waged an internal revolution in Peru. They began in Ayacucho (a Peruvian state), as a committee in the pro-Chinese faction of the communist party. By recruiting university students who became teachers in rural areas, the Sendero Luminoso were able to indoctrinate the peasants. This allowed them to create an insurgent group consisting mainly of rural Peruvian youth.

Through a combination of force and persuasion, the Sendero Luminoso wrested authority from the local government in many rural areas. In the cities, the Sendero Luminoso engaged in terrorist activity. The Peruvian government eventually deployed their army to try to arrest the spreading power of the Sendero Luminoso, resulting in many civilian casualties in what were described as "peasant massacres." Since it was often unknown whether a person was a senderista, the army's initial strategy was simply to kill anyone who was suspected. Thus, there were many casualties attributable to the Sendero Luminoso and many to the army's countermeasures (Ball et al. 2003; Sulmont February 2005). Apportioning this responsibility continues to be a controversial issue in Peruvian politics.

The Peruvian analysis had six available databases that had been collected by very different organizations: the Peruvian Truth and Reconciliation Commission (CVR[4]), the Defender of the People (DP), the National Coalition of Human Rights, the Agricultural Development Center, the Human Rights Commission, and the International Committee of the Red Cross. These last four datasets were ultimately combined to produce the summary statistics in Table 2 (see Ball et al. 2003, for a description of each of these datasets).

Table 2. Recorded number of killings across datasets in Peru

		Other databases			
		Yes		No	
		DP			
		Yes	No	Yes	No
CVR	Yes	1214	766	623	15,794
	No	1218	3354	1723	Y_{000} (?)

In Peru, MSE analyses were conducted using the log-linear model approach described in Section 2.4, incorporating stratification both in terms of geography and perpetrator group (see Ball et al. 2003, for stratification details). A single best model was selected within each stratum based on goodness of fit (in this case a χ^2 test was used; see Ball et al. 2003, for complete modeling details). These analyses resulted in an estimate of 69,280 deaths between 1980 and 2000, nearly three times the previously accepted estimate of 25,000 deaths often reported by human rights NGOs and newspapers (Ball et al. 2003). Additionally, the common local narrative of violence was that government forces had been the main perpetrators. MSE analyses revealed that violations committed by the Sendero Luminoso were disproportionately underreported, and that in fact this group was responsible for between 41% and 48% of killings and disappearances (Ball et al. 2003). Figure 2 shows the estimated total number of victims by perpetrator group and geographic location. We see that the estimated number of victims of the Sendero Luminoso (dark gray bars) is greater than the estimated number of victims of state agents (light gray bars) in every geographic region except Lima Callao. White bars represent the number killed by other groups or individuals.

3.3 Colombia

Colombia is burdened with ongoing unrest. Government forces, guerrilla groups, drug cartels, and paramilitaries are the actors in the longest-running internal armed conflict in South America. The International Center for Transitional Justice describes the violence as fueled by "struggles to control the population, land, natural resources, political power, and drug markets,"[5] while Amnesty International states that there is little agreement regarding the underlying causes of the conflict (Amnesty International 2008).

Data on conflict-related mortality is currently being collected by many different groups, despite the difficulties posed by unsettled and often insecure circumstances. Many groups, both within Colombia and the broader international community, have a declared interest in measuring the number of homicides, or lethal violations, occurring in Colombia. More specifically, patterns of lethal violations over time and space must be assessed to evaluate claims that demobilization of paramilitaries and amnesty laws have led to decreased violence.

[2] *http://www.riosmontt-trial.org/trial-background/*
[3] Personal correspondence with Patrick Ball.
[4] Acronym reflects the project's name in Spanish.
[5] "Background: Demobilization, Justice and Peace Law, and Other Initiatives." Available at *http://ictj.org/our-work/regions-and-countries/colombia*; accessed 1/5/11.

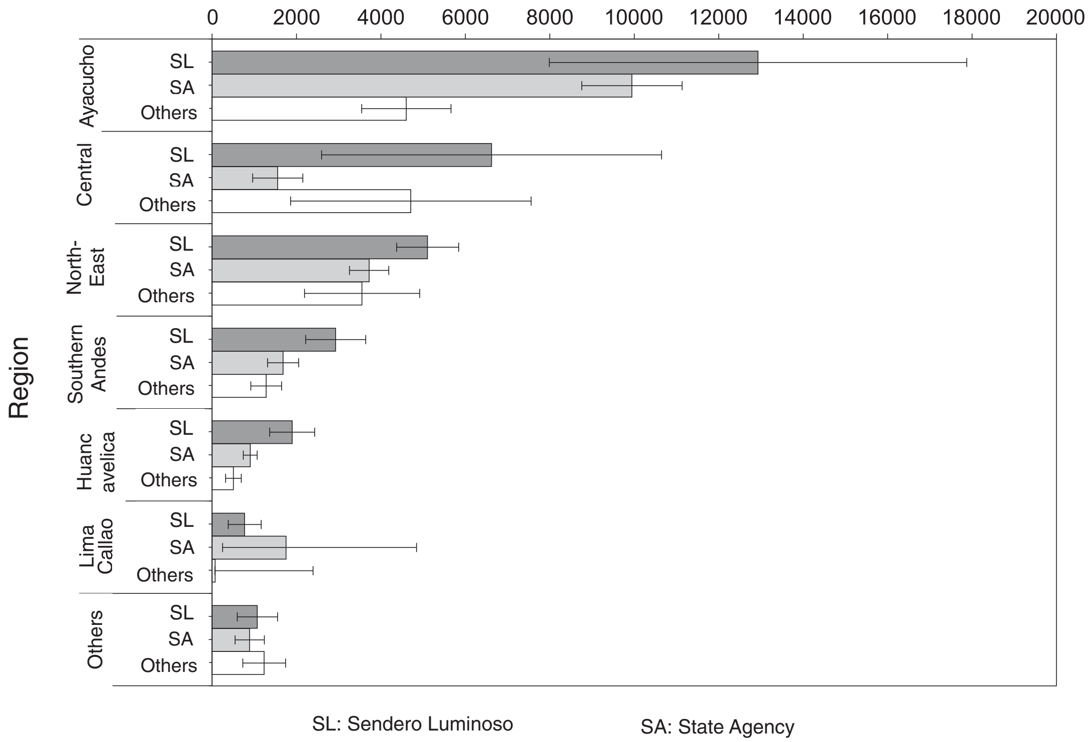

Figure 2. Estimated number of victims by region and perpetrator (modified from Ball et al. 2003).

Compared to the previous examples, the statistical analysis in Colombia had access to the largest number and most diverse kinds of datasets. The 15 data sources used for the Colombian study come from state agencies (including security, forensic, and judicial bodies) and from civil society organizations.

The large amount of data available to the Colombia project allowed the creation of many models that were similarly plausible but that led to substantially different estimates. This motivated the use of model averaging, as described in Section 2.5. In this framework, we write $\boldsymbol{Y} \sim \text{Multinomial}(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = \{\boldsymbol{\theta}_{C_1}, \boldsymbol{\theta}_{C_2}, \ldots, \boldsymbol{\theta}_{C_J}\}$ is the collection of probabilities of a violation occurring in each cell of the multinomial distribution. Each $\boldsymbol{\theta}_{C_j}$ is the vector of capture probabilities for clique C_j in the decomposable graphical model describing the dependence structure of the systems. In the Bayesian framework, we let each $\boldsymbol{\theta}_{C_j} \sim \text{Dirichlet}(\boldsymbol{\delta})$, which is the conjugate prior for the multinomial likelihood and allows for easy marginalization over the nuisance parameters, $\boldsymbol{\theta}$, for posterior inference about N. Thus, one can obtain closed-form expressions for the posterior probability, $P(N|\boldsymbol{Y})$, free of all other unknown model parameters (for a detailed description of the use of graphical models to represent list dependence, see Madigan, York, and Allard 1995, and for an introduction to graphical models generally, see Lauritzen 1996).

The major findings of the research in Colombia were that (1) even with access to 15 datasets, a statistically significant number of events are unreported (i.e., our 95% posterior credible interval for the total number of undocumented events did not include zero), (2) the pattern of unreported events varies over time and space, and (3) descriptive comparisons of the 15 datasets revealed that even similar agencies document different cases. For example, the Office of the Vice Presidency and the National Police both had access to much of the same raw data but produced very different lists. It is important to note that this last point is not a criticism of the data collection mechanisms of the Vice Presidency or National Police (or any of the other organizations that provided datasets for the Colombia analysis) but rather a key reality that different organizations capture different portions of the universe of interest, even when contextual knowledge leads us to expect similar samples.

4. MODEL-BASED SIMULATION

The three case studies provide examples of important applications for which the traditional assumptions underlying the classic two-system capture-recapture methodology do not hold. Additionally, the case studies show the evolving complexity of MSE analysis, as practical and political constraints drive researchers to use larger numbers of databases with complex dependency structures. Finally, as noted in Section 2.6, there is no way to guarantee that estimates are, in some sense, conservative. To address these concerns, we present an example of the use of model-based simulation to assess overall uncertainty and the impact of the many potential sources of uncertainty on MSE estimates.

To illustrate how simulation can be used to test the performance of MSE under violations of assumptions, consider an example analysis based upon, but not identical to, the Guatemalan case study. In this simulation, we begin with a subset of the records from the Guatemala case study summarized in Table 1. In the following description of the simulation procedure, this fixed subset of records is called the "original records." We repeat the following procedure 1000 times for each scenario.

1. Create an artificial population of violations by using all of the original records that appear in the Guatemala dataset plus K additional records.
   (a) If $K = 0$, use only the original records.
   (b) If $K > 0$, sample (nonuniformly) with replacement from the original set of records. Let the probability of sampling vary with the ethnicity variable so that certain ethnicities are more likely to appear in population of simulated violations.
2. Create synthetic lists by sampling from the artificial population created in Step 1. Sample records for List i with marginal probability p_i. Probabilities for individual records vary by sex, ethnicity, etc.
3. Stratify the samples based on geographic region and perform MSE using the R package `Rcapture`, selecting the single best-fitting model based on BIC.

We note that although we use geographic strata for estimation and different sampling probabilities by ethnicity, we mean this to be a generic simulation example. Any variables could have been used for stratification or creating inhomogeneous list probabilities. This is not meant to be a reanalysis of the Guatemala case study. Rather, we use the Guatemala data as inspiration for creating a plausible scenario.

The results of this experiment are shown in Table 3. For each of the seven scenarios presented, simulation parameters, K and $p = \{p_1, p_2, p_3\}$ are given as well as the range of unobserved counts N_{000} across the one thousand replicates of the simulation. The ground truth column shows the true number of records in the (artificial) population for each stratum. The simulations varied the coverage rate of the lists, such that the number of unobserved records varied from only a few hundred to several thousand. For each scenario, we report the mean of the estimates from the 1000 replicates and the 95% simulation interval (the 0.025 and 0.975 quantiles of the estimates). Simulations 1 through 4 maintain the odds of sampling records in the two simulated ethnic groups, three geographic regions, and sexes to be the same as the odds that appear in the real lists. This produces between-list dependencies and heterogeneous capture probabilities within list. In simulations 5, 6, and 7 (shown with an asterisk), list two collected nonindigenous records from region B at a higher rate than appeared in the real list, creating substantial capture heterogeneity by ethnicity in region B. In one of these cases (simulation 7), stratum B had very wide 95% simulation intervals. This hints at the fact that large levels of uncertainty may imply drastic violations of assumptions.

Overall, MSE performs robustly, although it tends to underestimate (with one exception) within this narrow set of simulation parameters. These scenarios demonstrate MSE performing well with relatively low capture probabilities (in some scenarios each list included between one-quarter and one-third of the total population) and in the presence of heterogeneous capture probabilities. This is indicated by the relatively small differences between the "true" population size and the average estimate across the simulations in most strata and most scenarios. Additionally, most simulation intervals contain the "true" population size—only 3 of the 21 intervals do not contain the true value. If an interval has 95% frequentist covarage, then we would expect exactly three of the 21 intervals to fail to cover the true value 6.6% of the time. Note that while it seems intuitive that these simulation intervals should have approximately 95% coverage, they should be interpreted with care, as we have not formally established their frequentist coverage properties.

These simulations explore only a small portion of the many steps involved in MSE analyses. A full-scale analysis, which is outside of the scope of the work presented here, would also model the deduplication and matching process by altering some

Table 3. Simulation results

#	Sim. details		Region	Ground truth	Mean	95% Sim. interval
1	N_{000}	[200, 297]	A	1012	986	(939, 1021)
	K	0	B	1263	1237	(1179, 1274)
	p	{0.29, 0.25, 0.85}	C	341	346	(288, 476)
2	N_{000}	[458, 580]	A	1012	953	(859, 1122)
	K	0	B	1263	1203	(1091, 1503)
	p	{0.29, 0.25, 0.65}	C	341	320	(237, 377)
3	N_{000}	[707, 865]	A	1012	874	(763, 1014)
	K	0	B	1263	1148	(983, 1207)
	p	{0.29, 0.25, 0.45}	C	341	250	(187, 263)
4	N_{000}	[957, 1053]	A	1012	773	(677, 1043)
	K	0	B	1263	987	(845, 1280)
	p	{0.29, 0.25, 0.25}	C	341	197	(149, 393)
5*	N_{000}	[457, 578]	A	1012	953	(859, 1122)
	K	0	B	1263	1203	(1091, 1503)
	p	{0.29, 0.25, 0.65}	C	341	320	(237, 377)
6*	N_{000}	[4316, 4696]	A	10,429	9548	(8480, 10,234)
	K	23,390	B	12,357	12,161	(11,839, 12,970)
	p	{0.65, 0.29, 0.25}	C	3322	2945	(2312, 3531)
7*	N_{000}	[11,410, 11,910]	A	10,429	7231	(5634, 11,102)
	K	23,390	B	12,357	12,371	(10,002, 21,204)
	p	{0.15, 0.29, 0.25}	C	3322	1534	(1040, 3603)

fields in historically plausible ways. For example, assigning a new first name from the core list of all first names and regenerating the location and date with probabilities corresponding to known violation intensities in different regions at different times. A full-scale analysis would also examine this step in the methodology by varying the γ parameter that determines whether two records are considered a match. Finally, one would also need to explicitly include falsified records and conduct a sensitivity analysis, varying the number of falsified records. Despite these omissions, we find these results promising regarding the validity of the Guatemalan results in that 18 of 21 of the simulation intervals contained the true number of deaths despite assumption violations that we believe mimic those that could be seen in the real application.

5. CONCLUSIONS

Using three case studies, we have demonstrated the application of MSE methods in human rights studies. In Guatemala, statistical analysis supported a claim of genocide.[6] In Peru, results of analysis with MSE indicated that Sendero Luminoso was responsible for a plurality of atrocities (46% compared to 30% of violations committed by agents of the Peruvian state) (Ball et al. 2003). And in Colombia, new MSE methodology found a much larger number of victims than expected, and exposed gaps in the record collection systems. The indispensable role of statistics in the field of human rights—and in other historically nonquantitative fields—should not be underestimated.

We have also illustrated how, in each of those three case studies, critical assumptions that underlie the original capture-recapture methodology are violated. Given the political controversy that surrounds such research—for example, as occurred in Peru following the announcement of the study results—it is important to improve MSE methodology so that it can provide more defensible estimates.

In this article, we argued that some classical assumptions are reasonable (e.g., that the population is closed), whereas others are problematic (i.e., homogeneous catchability and perfect matching). However, there is a great deal of domain knowledge that can be used to create detailed models of the kinds of assumption violations that occur, such as mismatch errors, underreporting of rural violations, or coverage bias in lists maintained by certain groups.

This article argues that one can improve understanding of the uncertainty in MSE methods through model-based simulation. These simulations are tuned to specific applications, and explore the impact of different degrees of violation on the ability of MSE to recover the truth. One output is a simulation confidence interval, whose coverage probability can be estimated since the true values used in the simulation is known. A second output is a better understanding of the kinds of biases that specific assumption violations introduce when they interact with the context of a given application.

This approach was tested on a subset of data from the Guatemalan case study, for some simple models of assumption violations. The main finding was that the MSE estimates were robust, and that considerable erosion of standard assumptions did not seriously undermine the ability of MSE to produce accurate estimates and confidence intervals with appropriate coverage. A second finding was that, depending on which assumptions were violated, the MSE inferences could overestimate or underestimate the truth; that is, there is no systematic way to produce an MSE procedure that is conservative when the assumptions are questionable.

[Received June 2012. Revised May 2013.]

[6] More precisely, a legal argument that the Army committed acts of genocide against certain Mayan groups in Guatemala (Ball 2000).

REFERENCES

Amnesty International (2008), *"Leave Us in Peace!" Targeting Civilians in Colombia's Internal Armed Conflict*, London: Amnesty International Secretariat. [196]

Ball, P. (1999), *Making the Case: Investigating Large Scale Human Rights Violations Using Information Systems and Data Analysis* (Vol. 12), Washington, DC: American Academy for the Advancement of Science. [195]

——— (2000), "The Guatemalan Commission for Historical Clarification: Intersample Analysis," in *Making the Case: Investigating Large Scale Human Rights Violations Using Information Systems and Data Analysis*, eds. P. Ball, H. Spirer, and L. Spirer, Washington, DC: American Association for the Advancement of Science, chap. 11. Available at *http://shr.aaas.org/mtc/chap11.html*. [191]

——— (2001), "Making the Case: The Role of Statistics in Human Rights Reporting," *Statistical Journal of the United Nations Economic Commission for Europe*, 18, 163–174. [195]

Ball, P., Asher, J., Sulmont, D., and Manrique, D. (2003), *How Many Peruvians Have Died? An Estimate of the Total Number of Victims Killed or Disappeared in the Armed Internal Conflict Between 1980 and 2000*, Report to the Peruvian Commission for Truth and Justice (CVR), Washington, DC: American Association for the Advancement of Science. [191,192,193,196,199]

Ball, P., Betts, W., Scheuren, F., Dudukovich, J., and Asher, J. (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. [191,195]

Ball, P., Kobrak, P., and Spirer, H. (1999), *State Violence in Guatemala, 1960–1996: A Quantitative Reflection*, Washington, DC: American Association for the Advancement of Science. [195]

Basu, S., and Ebrahimi, N. (2001), "Bayesian Capture-Recapture Methods for Error Detection and Estimation of Population Size: Heterogeneity and Dependence," *Biometrika*, 88, 269–279. [193]

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975), *Discrete Multivariate Analysis: Theory and Practice*, Cambridge, MA: MIT Press. [192,193]

Brunborg, H., Lynstad, T. H., and Urdal, H. (2003), "Accounting for Genocide: How Many Were Killed in Srebrenica?" *European Journal of Population*, 19, 229–248. [191]

Cairns, K. L., Woodruff, B. A., Myatt, M., Bartlett, L., Goldberg, H., and Roberts, L. (2009), "Cross-Sectional Survey Methods to Assess Retrospectively Mortality in Humanitarian Emergencies," *Disasters*, 33, 503–521. [191]

Chao, A. (1987), "Estimating the Population Size for Capture-Recapture Data With Unequal Catchability," *Biometrics*, 43, 783–791. [193]

Chao, A., Lee, S.-M., and Jeng, S.-L. (1992), "Estimating Population Size for Capture-Recapture Data When Capture Probabilities Vary by Time and Individual Animal," *Biometrics*, 48, 201–216. [193]

Cormack, R. M. (1989), "Log-Linear Models for Capture-Recapture," *Biometrics*, 45, 395–413. [193]

Coull, B. A., and Agresti, A. (1999), "The Use of Mixed Logit Models to Reflect Heterogeneity in Capture-Recapture Studies," *Biometrics*, 55, 294–301. [193]

Darroch, J., Fienberg, S., Glonek, G., and Junker, B. (1993), "A Three-Sample Multiple-Recapture Approach to Census Population Estimation With Heterogeneous Catchability," *Journal of the American Statistical Association*, 88, 1137–1148. [191,193]

Davenport, C., and Ball, P. (2002), "Views to a Kill: Exploring the Implications of Source Selection in the Case of Guatemalan State Terror, 1977–1996," *Journal of Conflict Resolution*, 46, 427–449. [191,192]

Elmagarmid, A., Ipeirotis, P., and Verykios, V. (2007), "Duplicate Record Detection: A Survey," *IEEE Transactions on Knowledge and Data Engineering*, 19, 1–16. [194]

Farmer, P. (2005), *Pathologies of Power: Health, Human Rights, and the New War on the Poor*, Berkeley, CA: University of California Press. [191]

Fellegi, I., and Sunter, A. (1969), "A Theory for Record Linkage," *Journal of the American Statistical Association*, 64, 1183–1210. [194]

Fienberg, S. (1972), "The Multiple Recapture Census for Closed Populations and Incomplete 2^k Contingency Tables," *Biometrika*, 59, 591–603. [193]

Fienberg, S., Johnson, M., and Junker, B. (1999), "Classical Multilevel and Bayesian Approaches to Population Size Estimation Using Multiple Lists," *Journal of the Royal Statistical Society*, 162, 383–405. [193]

Fienberg, S. E., and Manrique-Vallier, D. (2009), "Integrated Methodology for Multiple Systems Estimation and Record Linkage Using a Missing Data Formulation," *Advances in Statistical Analysis*, 93, 49–60. [194]

Guzmán, D. (2011), "Speaking Stats to Justice: Expert Testimony in a Guatemalan Human Rights Trial Based on Statistical Sampling," *Chance*. Available at *http://chance.amstat.org/2011/09/guatemala/* [191]

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009), "The WEKA Data Mining Software: An Update," *SIGKDD Explorations*, 11, 10–18. [194]

Hook, E., and Regal, R. (2000), "Accuracy of Alternative Approaches to Capture-Recapture Estimates of Disease Frequency: Internal Validity Analysis of Data From Five Sources," *American Journal of Epidemiology*, 152, 771–779. [191]

Huggins, R. (2001), "A Note on the Difficulties Associated With the Analysis of Capture-Recapture Experiments With Heterogeneous Capture Probabilities," *Statistics and Probability Letters*, 54, 147–152. [193]

International Working Group for Disease Monitoring and Forecasting. (1995a), "Capture-Recapture and Multiple-Record Systems Estimation I: History and Theoretical Development," *American Journal of Epidemiology*, 142, 1047–1058. [191]

——— (1995b), "Capture-Recapture and Multiple-Record Systems Estimation II: Applications in Human Diseases," *American Journal of Epidemiology*, 142, 1059–1068. [191]

King, R., and Brooks, S. P. (2008), "On the Bayesian Estimation of a Closed Population Size in the Presence of Heterogeneity and Model Uncertainty," *Biometrics*, 64, 816–824. [193]

Lauritzen, S. L. (1996), *Graphical Models*, Oxford: Oxford University Press. [197]

Lincoln, F. (1930), *Calculating Waterfowl Abundance on the Basis of Banding Returns*, Washington, DC: U.S. Department of Agriculture. [191]

Lum, K., Price, M., Guberek, T., and Ball, P. (2010), "Measuring Elusive Populations With Bayesian Model Averaging for Multiple Systems Estimation: A Case Study on Lethal Violations in Casanare, 1998–2007," *Statistics, Politics, and Policy*, 1, DOI: 10.2202/2151-7509.1005. [191,193]

Madigan, D., York, J., and Allard, D. (1995), "Bayesian Graphical Models for Discrete Data," *International Statistical Review/Revue Internationale de Statistique*, 63, 215–232. [193,197]

Manrique-Vallier, D., and Fienberg, S. E. (2008), "Population Size Estimation Using Individual Level Mixture Models," *Biometrical Journal*, 50, 1051–1063. [193]

Marks, E. S., Seltzer, W., and Krótki, K. J. (1974), *Population Growth Estimation: A Handbook of Vital Statistics Measurement*, New York: The Population Council. [191,195]

Peterson, C. (1896), "The Yearly Immigration of Young Plaice Into the Limfjord From the German Sea," *Report of the Danish Biological Station (1895)*, 6, 5–84. [191,192]

Phillips, L. (1990), "Hanging on the Metaphone," *Computer Language Magazine*, 8, 39–44. [194]

Raftery, A. E. (1995), "Bayesian Model Selection in Social Research," *Sociological Methodology*, 25, 111–163. [193]

Roberts, B., Morgan, O. W., Sultani, M. G., Nyasulu, P., Rwebangila, S., Myatt, M., Sondorp, E., Chandramohan, D., and Checchi, F. (2010), "A New Method to Estimate Mortality in Crisis-Affected and Resource-Poor Setting: Validation Study," *International Journal of Epidemiology*, 39, 1584–1596. [191]

Seber, G. (1965), "A Note on the Multiple Recapture Census," *Biometrika*, 52, 249–259. [191]

Seber, G. A. F., Huakau, J. T., and Simmons, D. (2000), "Capture-Recapture, Epidemiology, and List Mismatches: Two Lists," *Biometrics*, 56, 1227–1232. [191]

Sekar, C. C., and Deming, E. W. (1949), "On a Method of Estimating Birth and Death Rates and the Extent of Registration," *Journal of the American Statistical Association*, 44, 101–115. [191]

Silva, R., and Ball, P. (2006), *The Profile of Human Rights Violations in Timor-Leste, 1974–1999*, Report by the Benetech Human Rights Data Analysis Group to the Commission on Reception, Truth and Reconciliation (CAVR), Palo Alto, CA: The Benetech Initiative. [191]

Silva, R., Klingner, J., and Weikart, S. (2010), "Measuring Lethal Counterinsurgency Violence in Amritsar District, India Using a Referral-Based Sampling Technique," in *JSM Proceedings, Social Statistics Section*. Alexandria, VA: American Statistical Association. [194]

Sulmont, D. (2005), *The Peruvian Truth Commission (Comisión de la Verdad y Reconciliación-CVR)*, unpublished notes for the Conference at the Latin American Centre, University of Oxford. [196]

Tancredi, A., and Liseo, B. (2011), "A Hierarchical Bayesian Approach to Record Linkage and Population Size Problems," *Annals of Applied Statistics*, 5, 1553–1585. [194]

Turner, B. (2006), *Vulnerability and Human Rights*, University Park, PA: Penn State Press. [191]

Wittes, J. T., Colton, T., and Sidel, V. W. (1974), "Capture-Recapture Methods for Assessing the Completeness of Case Ascertainment When Using Multiple Information Sources," *Journal of Chronic Disease*, 27, 25–36. [191]

Wittes, J., and Sidel, V. W. (1968), "A Generalization of the Simple Capture-Recapture Model With Applications to Epidemiological Research," *Journal of Chronic Disease*, 21, 287–301. [191]

Woodbury, M. A., Clive, J., and Arthur Garson, J. (1978), "Mathematical Typology: A Grade of Membership Technique for Obtaining Disease Definition," *Computers and Biomedical Research*, 11, 277–298. [193]

York, J. C., and Madigan, D. (July 1992), "Bayesian Methods for Estimating the Size of a Closed Population," Technical Report 234, University of Washington. [193]

Zwierzchowski, J., and Tabeau, E. (2010), "The 1992–95 War in Bosnia and Herzegovina: Census-Based Multiple System Estimation of Casualties' Undercount," in The Households in Conflict Network (HiCN) and The German Institute for Economic Research (DIW Berlin) *Proceedings of the International Research Workshop on "The Global Costs of Conflict."* [191]