January 2015

Counting Uncounted Gunshot Injuries: A Capture-recapture Study Of People Minding Their Own Business

Zev Balsen
Yale School of Medicine, zev.balsen@yale.edu

Counting Uncounted Gunshot Injuries:
A Capture-Recapture Study of People Minding Their Own Business

A Thesis Submitted to the
Yale University School of Medicine
In Partial Fulfillment of the Requirements for the
Degree of Doctor of Medicine

Advisor: Lori Post, PhD; Department of Emergency Medicine

by

Zev J.C. Balsen

2015

Abstract

Objectives. To apply a novel statistical method to create a comprehensive estimate of incidence of firearm related injuries.

Methods. A database of firearms injuries in New Haven, Connecticut, during a five-month period was created with records from law enforcement, emergency departments, emergency medical services (EMS), news media, and the medical examiner. The overlap of these various sources was operationalized in a capture-recapture model to generate an estimate of uncounted firearms injuries, and log linear modeling was used to control for positive and negative dependencies.

Results. The combined data sources revealed 49 firearms injuries occurring during the study period within our defined geographical area. No single source recorded more than 43 of these injuries. Log-linear capture recapture methods estimated that the actual number of injuries was 49.7 (95% CI 49-52.3).

Conclusions. No single source reaches complete case ascertainment for firearms injures. Combining multiple sources improves the estimate of injury incidence, but still results in an undercount. Log-linear capture-recapture methods can be used to improve the estimate of firearms injuries.

Acknowledgements

This work could not have been completed without the hard work of a number of individuals. Thank you to Brandy Dionne and Pilar Ciravolo from the Yale Human Research Protection Program for their endless patience. Thank you to Desmond Walker for his assistance with obtaining data from Yale's EMR. Thank you to American Medical Response-New Haven, Yale-New Haven Hospital, the Connecticut Medical Examiner, and the New Haven Police Department for graciously sharing their records, without which we could not have undertaken this research. Thank you to Christal Esposito, Douglas Barber, Vanessa Kuhlor, and Michelle Wu for help collecting data and establishing the trajectory of the project. Thank you to David Cone and Liudvikas Jagminas for helping to access hospital and pre-hospital data. Thank you to Alexei Nelayev for advice and assistance regarding the Institutional Review Board. Thank you to Richard Spano for helping to shape the project, for contributing data and knowledge regarding the police department, and for many discussions along the way that guided the development of the research. Finally, thank you to Lori Post for her wise and careful oversight of the project.

Table of Contents

Introduction

Guns are an important and enduring part of American history and culture. Their benefits to society include control of pests, hunting and recreational target shooting,(1) crime deterrence,(2) protection,(3, 4) and cultural value.(5) Despite these benefits, a small percentage of gun uses results in injury or death.

This is what we think we know about the scope of the problem: In the United States, the rate of homicide from firearms is 20 times higher than the combined rates of 22 of our peer countries.(6) Annually more than 100,000 Americans suffer a gunshot injury and one in three people in the United States knows someone who has been injured.(7) One out of five teenagers ages 14-17 reports that they have personally witnessed a shooting.(8-10) The death toll from firearms affects rural and urban youth equally, although urban youth are more likely to die from homicide and rural youth are more likely to die from suicide or accidental shootings.(11, 12) Financially, firearm injuries have a heavy impact on society, with the lifetime medical cost for all gun violence victims in the US estimated at $2.3 billion. Almost half that cost is borne by taxpayers.(13-16) Clearly, gunshot wounds have a significant impact on public health in the United States. Data like those above provide a basis upon which to formulate interventions and to determine how best to direct those interventions and to judge the outcome of interventions. However, what if these data are erroneous?

Unfortunately, our understanding of the scope of the problem and how to intervene has been retarded by policies and a dearth of funding for firearm research since 1996.(17) Thus, in order to design, implement, and evaluate interventions to improve public health in relation to gun use, we need to first focus on quantifying and qualifying the scope of gun injuries and deaths.

Most current data on firearm injuries and violence relies heavily on a single source: law enforcement reporting. Although this is intuitively an obvious source of data, previous research has shown us that law enforcement data may significantly underreport the prevalence of gun-related events.(18) Furthermore, most research has been narrow in scope. Gun research has been restricted to studies of firearms, excluding other types of guns, such as air rifles, that are almost indistinguishable in terms of appearance and lethality.(19) In addition, much research has been restricted to “injuries” and “violence,” which excludes effects of guns such as coercion without injury and unintentional or self-inflicted injuries that may not always qualify as “violence.”

The current study demonstrates a useful method for studying the incidence and prevalence of gun events. Since no single source is likely to capture all relevant incidents, we propose matching multiple sources of data—including law enforcement, emergency department, news media, EMS, and medical examiner data—in order to obtain a more precise estimate. Combining multiple sources is likely to result in an undercount due to incidents that go uncounted by all sources. We can correct undercount-using capture-recapture techniques first established in wildlife biology and ecology.

Some might intuitively suspect that victims of gunshot injuries would report to law enforcement or to healthcare providers who would then report the incident to law enforcement; there are numerous situations in which this does not occur. For example, although healthcare workers are mandated reporters,(20) in the hectic environment of the emergency department, some wounds may not be reported.(21) Furthermore, there are numerous reasons why a victim of a gunshot wound may not report to the emergency department. Some injuries that are immediately fatal may route patients directly to the morgue, circumventing the emergency department. Some wounds may be minor and the victim may decide not to seek medical attention. In other cases, even with a significant injury, a victim may decide not to report the injury for several reasons: They may be prevented from reporting it by the perpetrator; they may be uninsured and fear medical costs; or they may not wish the authorities to be aware of the injury or of their whereabouts. This type of incident has even been depicted in the media.(22)

We propose the use of a method known as capture-recapture (also known as multi-system estimation) to generate a more comprehensive estimate of prevalence and incidence of gun related events. Originally developed for wildlife biology, this method is now established in public health research(23, 24) and in the study of injury(25-27), but has not been applied to the study of firearms. It uses various assumptions to generate an estimate of the proportion of uncounted members of a population.(28)

Existing Data on Firearms Injuries and Incidents

In 2005, the National Research Council of the National Academies convened a committee with the express goal of improving research, information, and data on firearms in the United States.(29) They determined that, despite the extent of the injuries, in the United States there is no authoritative source to provide accurate, timely, and complete data on the incidence and details of firearms-related injuries and violence. Instead, what we know about injuries comes from a patchwork of sources such as institutional data that suffer from clinical bias or samples that suffer from undercounts. At the national level, data sources include the National Crime Victimization Survey, the General Social Survey, Uniform Crime Reports, and the Bureau of Alcohol Tobacco and Firearms. Additional data comes from academic studies. Each of these sources contribute to our overall understanding of gun events but were produced with its own specific goals in mind and therefore each has its own advantages and drawbacks.(29)

The National Crime Victimization Survey, for example, relies on victim self-reports.(29) Although it is considered the "gold standard" for measuring crime victimization, it suffers from numerous methodological problems, such as nonreporting, false reporting, and nonstandard definitions of events. The Uniform Crime Reports rely on voluntary submissions of data from local law enforcement departments. This data set brings a separate set of problems, including the fact that not all incidents come to law enforcement attention, individual departments decide whether or not to participate, and criminal charges may change depending on the goals of local authorities.(29) Compounding this issue is the possibility that cities with the greatest amounts of violence may be the least likely to submit complete data.

The National Incident-Based Reporting System is meant to replace Uniform Crime Reports and is administered by the FBI.(29) This system contains great potential to record detailed information on firearm injuries across the United States. However, since it was first proposed in 1985, only a small percentage of the United States has been covered by the system, and notably most large cities and urban areas are not covered. The National Violent Death Reporting System is another national data set that has been administered by the Centers for Disease Control since 2002. In addition to methodological problems, including incomplete reporting and participation, it is not useful for studying firearms injuries because, as the name implies, it is aimed at collecting useful data only on incidents that result in death.(29) Furthermore, only a fraction of states are participating in data submissions.

Seminal independent studies that set the stage for our current work were led by researchers such as Kellermann and Hemenway.(5, 21, 30-53) Their landmark studies quantified and qualified firearm injuries as opposed to just firearms fatalities. In their studies, they observed that prior studies had been based on single data sources (e.g. hospital records, medical examiners, police records, etc.) and that they were therefore seeing only part of the picture. For example, Kellerman et al. combined data sources in three cities across the United States: Memphis, Tennessee; Seattle, Washington; and Galveston, Texas. For each city, they collected data from four main sources: police reports, emergency medical services agencies, emergency departments records, and medical examiners' records.

A large amount of valuable data about the characteristics of incidents involving gunshot wounds was collected, including age and race of the victims, types of weapons involved, severity of injuries, and the contexts in which shootings occurred.(21, 54-61) They found that the rate of firearms injuries per 100,000 residents per year ranged from 54 in Seattle to 223 in Memphis. 88 percent of the injuries occurred during assaults and only 4 percent of injuries were unintentional. The vast majority—88 percent—involved handguns. Interestingly, over 50 percent of the victims who arrived at the emergency department were admitted. Disturbingly, they found that the ratio of nonfatal injuries to deaths was only 4.2 to 1.

Our work builds on these earlier studies from Hemmenway and Kellermann's line of research. Kellermann et al specifically defined a case as an injury that would "prompt emergency medical attention."(21) Victims who did not seek care or those who sought care from practitioners who were willing to treat without reporting are missing from this study. The authors specifically state that their figures should be considered conservative. Of special relevance to our current study, they observed that despite the fact that health care providers in all three cities were required by law to report all gunshot wounds to law enforcement, they were unable to find matching police reports for a full 9 percent of cases identified. They determined that this lack of matching was most likely not due to misfiled or misplaced records but to true lack of reporting in the fast-paced emergency department environment. Building upon these landmark studies, we operationalize the overlap of various data sources to estimate the number of uncounted gun events.

Firearms Research Moratorium

Research on gun events has been significantly hampered by a de facto moratorium on federal funding for the last eighteen years.(62) Around the same time that the above paper by Kellerman et al was published, Congress passed a bill that, beginning in 1997, made funds at the Centers for Disease Control unavailable for research related to firearms.(63) The National Rifle Association claimed credit for pressuring Congress to adopt these changes because they felt that research on gun violence was actually being used in attempts to effect restrictions on gun ownership rights.(64). Later, the funding restrictions were extended to other federal agencies and gun safety was further polarized(15, 53, 65-73). In 2009, a study by Branas et al(74) used a case-control method to address the question of whether carrying a firearm increased or decreased the risk of becoming a victim of a firearm assault. They reported that, although some defensive gun use does occur, on average possession of a firearm did not appear to have a protective effect on the individual. The National Rifle Association condemned this study(64) and shortly afterward, Congress extended the funding restrictions to cover the agency that had sponsored the study, the National Institutes of Health.(62, 75)

This freeze on funding persisted for nearly two decades. On January 16, 2013, in the wake of the Sandy Hook Elementary School Shooting, President Barack Obama took a step to reverse the dearth of research and issued a memorandum directing the Centers for Disease Control to resume research on the causes and prevention of firearms injuries.(76, 77) The executive action, however, did not provide funding for this research, although President Obama did include funds in his budget proposal for fiscal year 2015.(78) On May 21, 2014, Senator Edward J. Markey (D-Mass.) and Representative Carolyn Maloney (D-NY) introduced legislation to provide long-term funding of $10 million per year for six years to the Centers of Disease Control to study firearms injuries and prevention.(78) These changes may alleviate some of the information deficits regarding incidence and prevalence of firearms injuries. As this line of research moves forward, it is imperative to find a common ground and avoid further polarization of the topic. We propose that the conversation be reframed as “responsible gun ownership” rather than “control,” in the hopes that a common ground will result in both public and political will to promote gun safety efforts.(79-81)

Capture-Recapture: A Novel Approach to Studying Firearms Injuries

As noted above, a principle difficulty in studying the incidence and prevalence of firearms injuries is the lack of a single source of data that comprehensively catalogues gun incidents. Thus, in order to arrive at this information, it is necessary to match multiple sources of data. Doing this also enables us to apply the statistical technique of capture-recapture. This technique is especially useful in terms of quantifying illegal or invisible activities because it does not rely on complete case ascertainment, but rather statistically operationalizes the overlap between various data sources.(28, 82-87) Specifically, in the case of a complex human population, we propose to use a form of capture-recapture known as multiple-systems estimation and account for violations of the statistical assumptions of basic capture-recapture techniques.(88)

In its most basic form, the capture-recapture approach is a method used commonly in ecology to estimate the size of an unknown population.(89) The method can be understood quite simply with an example. Let us say we are attempting to estimate the population size (N) of a fishery stock in a small lake. We start by capturing a given number of fish, let us say 100. We then tag those fish and release them back into the lake. After allowing sufficient time for these fish to mix back in with the population as a whole, we catch a second sample of fish. In this second sample, the "recapture," let us say we have again caught 100 fish and we see that 10 out of those 100 are tagged. If certain assumptions are met, we can then assume that the proportion of marked fish in this second sample closely approximates the ratio of tagged fish in the whole population.(90) Since we know the total number of tagged fish in the population is 100 we can set up this equation:

$$\frac{100}{N} = \frac{10}{100}$$

Now, solving for N, we determine that the best estimate of the population size is 1,000 fish. In its general form, the equation is represented as:

$$\widehat{N} = \frac{Mn}{m}$$

where $\widehat{N}$ is the estimate of the population size. M is the number of animals initially captured and tagged. n is the size of the second sample and m is the number of marked animals in this sample such that m/n is the proportion of marked animals in the sample.(90)

Of course, even for this trivial example, a number of assumptions about the population of interest are required. These assumptions can be described in a number of ways, but the key components are(89, 91):

A. The population is closed. In other words, the population is constant during the study period. In the case of animal studies, this requires zero immigration, emigration, births, or deaths. An alternative way to view this assumption is as requiring that each individual in the population has a non-zero probability of being captured in each of the samples taken.

B. The individuals can be reliably identified. It is crucial that the investigator is able to reliably identify which individuals captured or not captured in one sample appear in another sample. In wildlife studies, this is often as simple as assuming that all marks, tags, or bands remain in place and legible and that they are correctly recorded. In studies with humans, which generally do not involve physical tags, it implies that each individual can be uniquely identified using characteristics such as name, date of birth, or identification number (e.g. social security number, medical record number, etc.).

C. Each individual has the same probability of being captured in a given sample. For example, in a study on fish, this assumes the absence of wise old fish who evade the hook.

D. The samples are independent. In a wildlife study, this assumption would require that individuals caught in the first sample do not learn to evade capture a second time. In a human study, in which the "captures" may actually be lists of names from various agencies, this assumption would require that appearance of an individual on one agency's list would not make the individual more likely to appear on another agency's list.

These assumptions must be met in the closed-population model. In most situations, it is not possible to meet these assumptions and more complex statistical controls are required.(88, 92, 93)

History of Capture-Recapture Methods

Capture-recapture, also known as dual-system estimation when applied to human populations(94), has been used for centuries to estimate sizes of unknown populations. The procedure has long been associated with the Danish fisheries scientist C.G. Johannes Petersen (1860-1920) who attempted to estimate populations of fish in the 19th century.(90) However, this approach had been used much earlier by Pierre-Simon Laplace to estimate the population of France in 1786,(90, 95) and even earlier by John Graunt to estimate the population of London and the effect of plague in the 1600s.(28, 90)

The method was expanded by Knut Dahl in 1917 to study trout populations in Norway.(90) A capture-recapture approach was first applied to non-fisheries wildlife ecology by F.C. Lincoln in 1930 when he used hunters' returns of leg bands on ducks to estimate duck populations in North America.(91) Since then, the method has become ubiquitous in ecology and employed with a great degree of sophistication in innumerable wildlife studies.(28, 96)

In principle, the capture-recapture method can be applied to any situation in which there are two partial lists of members of a population. In order to generalize the method, one simply replaces the idea of "being captured in a sample" with the idea of "being on a list."(96) Despite the fact that some of the earliest known applications of these methods involved human populations and disease in the 17th century,(90) for most of its history, the capture-recapture method was restricted primarily to wildlife studies. The first serious application to human health in the modern era was in 1949 by Sekar and Deming who applied the method to birth and death rates in India.(96, 97)

During the middle of the 20th century, there continued to be use of the method for analysis of census data. In fact, the United States Census Bureau has attempted to apply capture-recapture techniques to correct for undercounts.(98) These applications have been extremely contentious politically.(99) They have led to partisan debate because of the way that the census creates differential undercounts (i.e. minority citizens are more likely to be undercounted) and because applying capture-recapture techniques to correct counts can have effects on such matters as the allocation of seats in the House of Representatives.(100) In fact, the U.S. Supreme Court has even rejected use of statistical adjustments to the census for purposes of apportionment.(101)

Significant applications to human epidemiology have been less controversial but did not come until the landmark 1968 paper by Wittes and Sidel(102) who pointed out the connection with work in other fields and championed the method to the epidemiology community.(96) Crucially, this paper also highlighted the problem of positive and negative dependence between sources as is common in applications involving retrospective epidemiological data. They pointed out that when there is positive dependence between the sources, this may lead to an underestimate of the population size, and negative dependence may lead to an overestimate of population size.(23, 102) We will return to a discussion of how to mitigate the problem of dependence after first reviewing other applications of capture-recapture methods in epidemiology.

A 1995 paper authored by the International Working Group for Disease Monitoring and Forecasting (IWGDMF) reviewed the applications of capture-recapture in epidemiology.(103) They began by contrasting the approach to counting in ecology with the approach in epidemiology. Both fields are built on a foundation of population counts. However, whereas ecologists have long recognized that undercount is inherent in any monitoring system, epidemiology has labored under the assumption that accurate incidence and prevalence data depend on enumeration that achieves near perfection. In other words, the IWGDMF argued that in epidemiology we tend to assume that the number of missing cases is negligibly small. Of course, incomplete case ascertainment is inherent in all monitoring systems. Although at times it may be negligibly small, it is necessary to quantify it, and when appropriate, correct for this undercounting.

With the above argument as a background, the IWGDMF paper summarized the history of capture-recapture applications in epidemiology.(103) They noted that following the 1968 paper by Wittes and Sidel(102) and a 1969 paper by Lewis and Hassanein (104) on hospital infections, the method still did not catch on until the 1980s. Coinciding with increased access to statistical computing technology, they noted an increase in the application of the log-linear methods, which are necessary to control for the dependencies and violation of elementary capture-recapture assumptions in most human studies. They noted the use of the method in papers that could be group into six different disease groups: birth defects, cancer, drug use, infectious disease, insulin-dependent diabetes mellitus (IDDM), and injuries. At the time of their writing in 1995, they found only 6-10 papers in each of the first four of those categories. These papers covered topics such as spina bifida, fetal alcohol syndrome, breast cancer, heroin use, and sexually transmitted disease. Interestingly, they noted that in the case of IDDM, capture-recapture methods had become standard around the world. Almost all type 1 diabetes registries used the method to check their ascertainment. Surprisingly, and in stark contrast to the IDDM literature, they found only 4 papers from the injury literature that applied the method.

Since that 1995 IWGDMF review, the application of capture-recapture to the study of injury has expanded somewhat. For example, a PubMed search on December 13, 2014, including the terms “capture,” “recapture,” and “injuries” produced 55 results. Three of these results were from 2014. Many of the papers address workplace injuries or traffic injury. Specific examples related to injury include traffic accidents in Scotland(105), all-cause pediatric injuries in Pittsburgh(106), transportation-related injuries in Nicaragua(107), dog bites in Pennsylvania(25), workplace injuries(108, 109), intra-family violence(86), and intimate partner violence(27). However, the inappropriate application of capture-recapture methods to dual-system cases that violate the method's assumptions has also been criticized in the context of injury studies.(110)

Dealing with Dependence

As described above, classical methods of capture recapture rely on a number of assumptions. In human studies, assumption A regarding a closed population is generally possible to meet. Because humans all have unique identifiers (e.g. names, birth dates, social security numbers, etc.) it is generally not a problem to meet assumption B that individuals in different samples can be reliably matched. However, in some studies in developing countries where identifiers may not be as reliable, even assumption B can be a challenge.(23) Unfortunately, assumptions C and D regarding equal "catchability" of all subjects and independence of lists are generally not possible to meet in human studies.(23, 111) For example, if one of the agencies producing a list refers individuals to another agency producing a separate list, there will be positive dependence between lists.(111) Likewise, if one of the agencies is a help-giving organization, then individuals seeking help may be more likely to appear on their list and therefore not all individuals will be equally likely to be counted.(111)

To illustrate this point, consider the type of case covered in our study: If a victim suffers a gunshot wound and travels to the emergency department for treatment, the treating physician has an obligation to report the injury to the police department,(20) and the victim will therefore have both an emergency department record and a police record. This would be described as a positive dependency between the emergency department and the police department.(112) Conversely, a victim of a gunshot wound who dies in the field would be transported to the morgue rather than to the emergency department. This case would therefore register in medical examiner records but not in emergency department records. This would be considered negative dependency between the medical examiner and the emergency department.(86)

If estimation is attempted based on only two sources of data (i.e. dual system estimation), dependencies between data sources can interfere with the ability to generate an accurate estimate.(23) However, if we introduce additional samples using more than two data sources (i.e. multiple system estimation) we can control for positive and negative dependencies using log-linear modeling.(86, 88, 111, 113)

The log-linear model is the standard form of analysis for contingency tables and was first proposed and developed for use with capture-recapture models by Fienberg in 1972.(93, 111) The foundation and application of this technique have been well described by the IWGDMF(111) and summarized by Chao et al.(23) The approach begins with regarding the data as a contingency table with 2^t cells. In this case t is the number of lists and each cell is a combination of lists on which individuals can appear. We can see that the number of cells therefore increases exponentially as we add lists. To generate the table, the presence and absence of an individual on a list is symbolized by 1 and 0, respectively. We can illustrate this with an example in which there are three lists. If a given individual shows up on the first list but not on the other two, he would be recorded as [100] in the data table. If an individual shows up on all three lists, he would be coded as [111]. Thus, the seven possible ways an individual can be recorded are: [001]; [010]; [100]; [110]; [101]; [011]; and [111]. The uncounted individuals are represented by the cell [000], but of course this cell starts out empty because we do not yet know how many uncounted individuals are in the population. With this established, software is used to fit log-linear models to the seven cells that contain observations. The optimal model is usually then selected with the Akaike information criterion (AIC). With this model chosen, it can then be projected onto the cell of uncounted individuals, [000], to estimate the number of uncounted individuals.

It is important to note that no matter how well our data fit a log-linear model, we can only estimate the number of uncounted individuals if we make the assumption that the inclusion pattern of the uncounted individuals is similar to the pattern of individuals that we were able to observe.(88) In other words, our model is only valid if the uncounted individuals are similar to the counted individuals in terms of their probability of being counted.

The model works best if the population has homogenous probabilities of being counted. If in fact the population is heterogeneous, we can theoretically improve the model by first stratifying it into groups that may be more homogenous, such as race, gender, age, or criminal status.(88) Estimates of total count can then be modeled within each stratum.

As stated earlier, the methods described here have taken on an increasingly prominent role in epidemiology during the last several decades. They have been applied to a wide range of problems in injury prevention, but have not as yet been applied to the study of firearms injuries.

Statement of Purpose and Hypotheses

The purpose of this project is to briefly explain the history and politics of gun research, redefine "firearm violence" to "gun events," elucidate limitations in previous research, reframe the polarized literature and advocacy from Second Amendment rights versus strict gun control to a moderate framework of responsible gun ownership, and to investigate the application of a novel method to the epidemiology of gunshot injuries. We generate gun incident estimates that can inform the planning, execution, and evaluation of interventions to reduce the impacts of gunshot events on public health, the economy, and quality of life.

Specific Aims

The specific aims of this study are:

1. Generate an estimate of total gunshot injuries occurring in a defined geographical area during a defined period of time.
2. Identify characteristics of victims (e.g. gender, race, criminal background, etc.) that influence the probability of suffering a gunshot wound and, once an injury is suffered, the probability of being counted on the lists of various agencies by stratifying estimates.
3. Evaluate the availability and suitability of various data sources to build a statewide system to monitor trends in gunshot injuries.

Specific Hypotheses

The specific hypotheses of our study are:

1. No single data source will achieve universal case ascertainment of gun events
2. Simple matching and summation of data from multiple sources will result in an undercount because of incomplete case ascertainment.
3. Capture-recapture methods using log-linear modeling for multiple data sources will allow for an evaluation of the degree of ascertainment and generate ascertainment-corrected rates.

Methods

In this study, we set out to establish a method for more comprehensive estimation of the number of injuries due to firearms. We hypothesized that no single source of data would record all firearms injuries that occur. Police files would record many incidents. Not all incidents, however, would be recorded by them. For example, although the emergency department is mandated to report all treated gunshot wounds,(20) in the chaotic environment of the emergency department not all patients with suspected gunshot wounds will necessarily be reported to the police. Likewise, the emergency department will record many but not all gunshot wounds. Some victims may feel that their wounds do not need to be treated by a doctor or may avoid treatment because they fear being reported to the police. Others may be unable to reach a hospital or feel they are unable because of insurance status, lack of transportation, or duress. Finally, victims who are declared dead on the scene will bypass the emergency department. Thus we set out to first collect data from a variety of different sources in order to collect as many gunshot cases as possible. After collecting this data we excluded cases that did not meet our geographic criteria, temporal criteria, or case definition. We then matched cases based on personal identifiers and estimated uncounted cases using capture-recapture methods and log-linear modeling.

Institutional Review

Our study design, methods, and goals were evaluated by the Yale University Institutional Review Board and approval to execute the project was granted. Amendments to personnel and data collection requirements were submitted and received approval as needed.

Study Area

We restricted our study geographically to injuries that occurred within the City of New Haven, Connecticut. This is not synonymous with New Haven County, a larger geographical area that includes several neighboring towns. The City of New Haven is located on the Connecticut shoreline, approximately midway between New York City and Providence, Rhode Island. According to the 2010 United States Census the population of New Haven was 129,779.(114) Within that population, the median age was 30.4 years. 22.8% of the population was under 18 years old. The population was 48% male and 52% female. Caucasians accounted for 42.6% of the population. African Americans accounted for 35.4%. Regardless of race, 27.4% of the population identified as Hispanic or Latino. Regarding education, 81.3% had at least a high school education. 26.5% of the population was living under the poverty line.

New Haven was chosen for two reasons. First, it is a location in which a sufficient number of gunshot injuries(115) are expected to occur in order to perform meaningful statistical analyses. Second, there are only three hospital emergency departments in the city, the Saint Raphael emergency department and the Yale-New Haven pediatric and adult emergency departments. All three of these departments are run by the Yale-New Haven Hospital system and there are no other significant trauma centers in the area, so all gunshot injuries that came to the emergency department are available to us through the Yale-New Haven electronic medical record system.

Time Period

Temporally, we restricted our study to the period of August 1, 2013, though December 31, 2013. We expected that this would be a sufficient period of time in which enough gunshot injuries would occur to permit statistical analyses. Furthermore, the study period concluded long enough ago that we could reasonably expect most police department case records to have been completed and filed and therefore available for our analysis.

Case Definition

In order for an incident to be included, it needed to involve an injury due to discharge of a firearm and striking of the victim by the discharged projectile (bullet). Both fatal and non-fatal injuries were included. We excluded injuries due to nonfirearms, such as BB guns and nail guns. We also did not include injuries due to firearms that did not involve discharge of a projectile, such as being struck by a firearm ("pistol-whipping").

Data Sources

Emergency Departments (ED)

The Yale-New Haven Hospital (YNHH) system owns all major hospitals in the New Haven area. The hospital began in 1826 and built its first building with 13 beds in 1833.(116) Today the hospital has grown into a 1,541-bed private, nonprofit teaching hospital. The primary hospital campus at York Street includes a Level 1 trauma center and emergency department with over 100,000 patient visits per year, including 70,000 adults and 30,000 children.(117) In 2012, YNHH acquired the nearby Hospital of Saint Raphael,(116) which included another major emergency department for the city. Because of this acquisition, the records of both the main YNHH emergency department and the Hospital of Saint Raphael emergency department are accessible within a single electronic medical record system.

The YNHH system uses Epic (Epic Systems Corporation, Wisconsin, U.S.A) as its electronic medical record system. We queried the YNHH Epic database for records that included the term *gun* or *GSW* (a common abbreviation for "gunshot wound") in the fields for *Diagnosis*; *Chief Complaint*; *Reason for Visit*; *Arrival Complaint*; *Injury Type*; and *Weapon/Type of Assault*. Although, most cases should have been noted as gunshot wounds under diagnosis or chief complaint, we searched those additional fields in order to ensure capturing as many cases as possible, given the fact that data entry may be inconsistent when a critical gunshot victim arrives in the emergency department.

Police Department (NHPD)

The New Haven Police Department, headquartered at 1 Union Avenue, New Haven, serves the City of New Haven. Annually, the department responds to approximately 8,000 Part I crimes as defined by the F.B.I.'s Uniform Crime Reports.(118) This figure includes approximately 2,000 major violent crimes (murder, forcible rape, robbery, aggravated assault) and 6,500 property crimes.

We requested data for all incidents classified as *Shooting* or *Murder*, including both fatal and nonfatal events, in the New Haven Police Department database during our specified time period.

Emergency Medical Services (AMR New Haven)

American Medical Response (AMR) New Haven is the only transporting emergency medical services agency within the city of New Haven. They staff with both basic life support and advanced life support, responding to a total of 100,000 calls per year.(119)

AMR New Haven searched their database for all calls for *Penetrating Trauma*. They then excluded cases that were not related to gunshot injuries (e.g. stabbings, eye injuries, animal bites, etc.). The resulting data set included records for all calls for gunshot wounds that originated at an address within the city of New Haven during our study period. The AMR database also enabled us to crosscheck emergency department records to determine where shootings occurred for other cases to confirm if they did or did not occur within the limits of the City of New Haven.

Medical Examiner (ME)

The Office of the Chief Medical Examiner, located in Farmington, Connecticut, provides certification of the cause of death for deaths occurring in the state of Connecticut.(120) We requested data pertaining to all homicides and suicides occurring in New Haven County during our study period.

News Media

New Haven news is covered by two newspapers, the *New Haven Independent*(121) and the *New Haven Register*.(122) It is also regularly covered by the *Hartford Courant*,(123) a newspaper headquartered in Hartford, Connecticut. Local television news media include WTNH, an ABC affiliate,(124) and Fox CT.(125) News media can often provide rich details about the background of the incident and the parties involved. It has been previously used in a similar manner in conjunction with law enforcement records and medical examiner reports to study fatalities related to intimate partner violence.(126)

We searched Google, Google News Archive, and the websites of local newspapers for all results that included the term *New Haven* combined with the terms *shooting*, *gun*, *gunfire*, *gunshot*, or *firearm*.

Data Processing

After collecting data, we systematically examined each set to eliminate cases that did not meet the eligibility criteria for our study. Specifically, cases needed to occur within the City of New Haven, within the specified period of time, and involve a gunshot injury as defined above. Determining if the incident occurred with the specified time period was simple because each data set clearly specified when the incidents occurred. Determining if the case met our definition of a gunshot injury was also straightforward because each data set provided details regarding the incident and specified if a weapon other than a firearm (e.g. BB gun) was involved. Determinations about location of the incident presented more of a challenge. Data from the police department, medical examiner, and AMR New Haven specified exact locations, making it simple to determine if the incident met our study criteria. News media sources also often specified location. Emergency department data, in contrast, rarely included information specifying location of the incident. Therefore, in order to determine locations of incidents in the emergency department data, we relied on auxiliary information in the other data sources after matching. News media sources were especially helpful at this stage because they enabled us to search specifically for shooting incidents that appeared in the emergency department data but may not have occurred in New Haven and therefore were absent from the other data sets.

Record Linkage

After collecting data and removing cases that did not meet our inclusion criteria, we combined all data sets and linked records via a manual process of fuzzy matching. Some cases we were able to match with exact name identifiers. However, heterogeneity in spelling of names across data sets required us to accept variation in spelling of names, provided that the data matched in terms of date of the incident and date of birth of the victim. For example, if one data set identified an individual as John Smith and another identified him as Jonathon Smith, but the birth dates and incident dates matched, we would consider this to be a single individual.

Resolving Inconsistencies

During and following the record linkage process, we attempted to resolve inconsistencies that arose regarding names, ages, and race or ethnicity of the victims. For example, if two records seemed to require linkage, but there was a dramatic difference in the name of the individual in different data sets (e.g. “John” and “Tim” for the same individual) we attempted to determine the true name of the individual by searching news reports, police records, and ambulance records. We attempted to determine if the inconsistency were due to a clerical error by the originating agency, use of a false identification by the victim, or simple lack of information (e.g. due to a victim being unconscious and unable to provide identifying information).

Statistical Analysis

All analyses were conducted in R version 3.1.2 (R Foundation for Statistical Computing, Vienna, Austria).(127) R offers a variety of packages for the handling of capture-recapture problems. These packages are summarized in the R Project's Environmetrics Task View.(128) Relevant packages include *Rcapture*, *BBRecapture*, *marked*, *unmarked*, *mra, secr, SPACECAP, and RMark.* The basic package *marked* provides a general framework for capture-recapture data handling and analysis. *Rcapture* provides tools primarily for log-linear modeling of capture-recapture data. Several packages are aimed toward spatial analysis of capture-recapture data. These include *secr*, *SPACECAP*, and *unmarked*. The package *mra* fits specific open and closed population models, including Cormack-Jolly-Seber, Huggin, and Horvitz-Thompson. *BBRecapture* uses a Bayesian approach to fitting of models with behavioral covariates. *RMark* provides a command-line R interface for the separate program MARK. Of these packages, only *Rcapture* is specifically geared toward log-linear models.

We primarily used the package *Rcapture* to create log-linear models of our data. We also used the package *fuzzySim* and the function *splist1presabs* to convert our data from a list of observations into a presence-absence table compatible with *Rcapture*.

This table took the format “00010; 01100; etc” with 0 indicating absence of an individual in a given data set and 1 indicating presence. We explored our data using heterogeneity graphing as described by Baillargeon and Rivest.(129) We used the *closedp* function to model our data because, although new incidents occurred over the course of our study period, the population was effectively closed (i.e. no immigration or emigration) because all data sets were collected concurrently. After model generation, we used Akaike information criterion (AIC), Bayesian information criterion (BIC), and boxplots of the Pearson residuals of our models to choose the best fitted model. The BIC is similar to the better known AIC, and despite the name, the two methods do not differ in terms of Bayesian versus frequentist perspective, but in terms of the way that they penalize overfitting of the data.(130)

Division of Labor (required statement)

Conception and oversight of the study was completed by Lori Post. Elaboration of the study details and parameters was completed by Zev Balsen (ZB) and Lori Post via an iterative and collaborative process. Requests for data through EPIC and AMR were executed by ZB. Records were obtained from the New Haven Police Department by Richard Spano, Assistant Professor of Criminal Justice at University of New Haven in collaboration with ZB. Medical Examiner records were obtained by Lori Post. Data collection from news media sources was done by Doug Barber, Vanessa Kuhlor, and Michelle Wu under the direction and supervision of ZB. Data processing, matching, and statistical analyses were completed by ZB.

Results

The data collected from the five different sources are summarized here and in Table 1 shown below.

<u>Police Department (NHPD)</u>

The data from the NHPD included 46 records. One incident involved an injury from a BB gun and was eliminate from further analyses, leaving 45 records. Two of these records were duplicates and were removed, leaving 43 unique victims of gunshot injuries. A large proportion (86%) were male. There was a large range of ages from 15 years old to 77 years. The mean age was 29.3 years. Two victims were under 18. The vast majority (95%) of victims were described as black or African American. Only two victims were Hispanic. None were described as non-Hispanic white or unknown. Approximately one in five of the incidents had a fatal outcome. One additional event was described as "likely to become fatal" due to the severity of the injury.

<u>News Media</u>

We originally located 83 news reports of gunshot injuries in New Haven during our study period. After removing duplicates, we had recorded 35 gunshot injury victims. Seven of these individuals were not identified by name in the news sources. However, during matching, there was sufficient data regarding age, injured body part, address where shot, and gender to confidently match all of these individuals.

For one of the victims, however, it was unclear if she met the case definition. She self-presented to a police department in an apparent state of intoxication and reported having been shot. However, the law enforcement officers appreciated no injury or damage to her clothing. At the alleged address of the incident, police officers found no witnesses, shell casings, or other evidence that a shooting had occurred. Despite the ambiguity in this case, it remained in the data set because it had been reported as a shooting by the woman and by the news media. This same individual also appeared in the AMR data set and was classified as a shooting there as well.

The remainder of the incidents reported in the news media had similar characteristics to those recorded by the NHPD. 83% were male and 91% were black. All 9 fatal incidents were reported.

Emergency Medical Services (AMR New Haven)

The data from AMR New Haven included 40 records. One record was for a BB gun injury and was eliminated. Three records were duplicates and were eliminated. This left 36 victims recorded. Of these 36, three were unidentified. However, during matching, there was sufficient data regarding age, injured body part, address where shot, and gender to confidently match all three of these individuals. One individual gave a name that was recorded on the AMR records, but based on matching of other data elements—and a news article saying the man initially provided a false identification—he was eventually identified as a different individual for our further analyses

Of the final 36 records, the characteristics were similar to those identified by the NHPD. 83% were male. 92% were black. However, the AMR records did identify one victim as unknown race. The AMR records captured 8 of the 9 fatalities recorded by the police. Presumably, the 9^{th} fatality was unequivocally dead on the scene and no ambulance response was requested.

Medical Examiner (ME)

The data from the Office of the Medical Examiner included 188 records during our study period. Of these, only 9 met our criteria of involving gunshot injuries and taking place in the City of New Haven. Of course, all of these incidents were fatalities. These 9 incidents were the same 9 fatalities recorded by the NHPD data set. 100% were homicides; no firearm suicides were recorded in New Haven during the study period. Of these 9 homicide victims, 89% were male. 89% were black and 11% (1 victim) were Hispanic. There were no victims under 18 years of age. Age ranged from 18 to 36. The mean age was 24. Four of the incidents involved gunshots to the head. Five involved gunshots to the chest/trunk. Two were declared dead at the scene where they were shot. The remainder were declared dead at the hospital.

Emergency Departments (ED)

The data retrieved from the YNHH EMR included 72 records. More than half of these records, however, were eliminated because they did not meet our inclusion criteria regarding geographical location, date of incident, or mechanism of injury. Six records were eliminated initially because they did not involve firearms (2 were Taser injuries, 2 were from nail guns, and 2 were from BB guns). An additional 15 records were eliminated because they were at the emergency department in Bridgeport, CT, and were highly unlikely to have occurred in New Haven. Two more records from patients who were transported to New Haven from Wallingford, CT, were eliminated. Following these initial removals, 7 more records were deleted after being found to be duplicates. A final 11 records were eliminated after cross-checking them with records from AMR-New Haven and with internet searches of the news media to determine that they had been shot outside of New Haven.

The data trimming described above reduced the total number of records to 31. Of these 31, there were 3 for which we had insufficient data to determine if they did or did not meet our inclusion criteria. This is because these records had no overlap with our other four richer data sets in order to determine exact location, dates, and circumstances. We chose to analyze and present both the complete data as well as the reduced set with the 3 equivocal cases excluded. For the complete data, there were interesting similarities and differences between the ED data and the other four data sets. Again, the majority were male (81%). Age range, mean, and number of pediatric patients were similar to other data sets. Interestingly, this was the only data set that included any victims identified as non-Hispanic white. A total of 3 victims (10%) were identified as non-Hispanic white. One of these victims, however, was also documented in the NHPD data set, but he was identified as black there. Despite the increased number of white victims, the percentage of black victims remained similar at 87%. This was due to there only being one victim identified as Hispanic. The number of fatal cases was dramatically lower than in the other data sets. Only 1 patient (3%) was a fatality. We will address the implications of this finding in the discussion section below.

TABLE 1. CHARACTERISTICS OF VICTIMS OF FIREARM INJURIES

RECORDED BY FIVE SOURCE IN NEW HAVEN, CONNECTICUT

AUGUST 1, 2013 – DECEMBER 31, 2013

	DATA SET[1]					
VARIABLE	**NHPD**	**News**	**AMR**	**ME**	**ED[3]**	**ED[4]**
Total victims	43	35	36	9	28	31
Unique victims[2]	0	0	0	0	1	4
Gender — no. (%)						
Male	37 (86)	29 (83)	30 (83)	8 (89)	22 (79)	25 (81)
Female	6 (14)	6 (17)	6 (17)	1 (11)	6 (21)	6 (19)
Age						
Range — yrs.	15 – 77	15 – 77	15 – 53	18 – 36	15 – 77	15 – 77
Mean — yrs.	29.3	29.2	27.9	24	30.3	29.8
< 18 — no. (%)	2 (5)	2 (6)	2 (6)	0	1 (4)	1 (3)
Race/ethnicity — no. (%)						
Non-Hispanic white	0	0	0	0	2 (7)	3 (10)
Black	41 (95)	32 (91)	33 (92)	8 (89)	26 (93)	27 (87)
Hispanic	2 (5)	2 (6)	2 (6)	1 (11)	0	1 (3)
Other or unknown	0	1 (3)	1 (3)	0	0	0
Fatalities — no. (%)	9 (21)	9 (26)	8 (22)	9 (100)	1 (4)	1 (3)

[1]Data Sets: NHPD = New Haven Police Department; News = news media source; AMR = American Medical Response-New Haven; ME = Medical Examiner; ED = YNHH emergency departments.

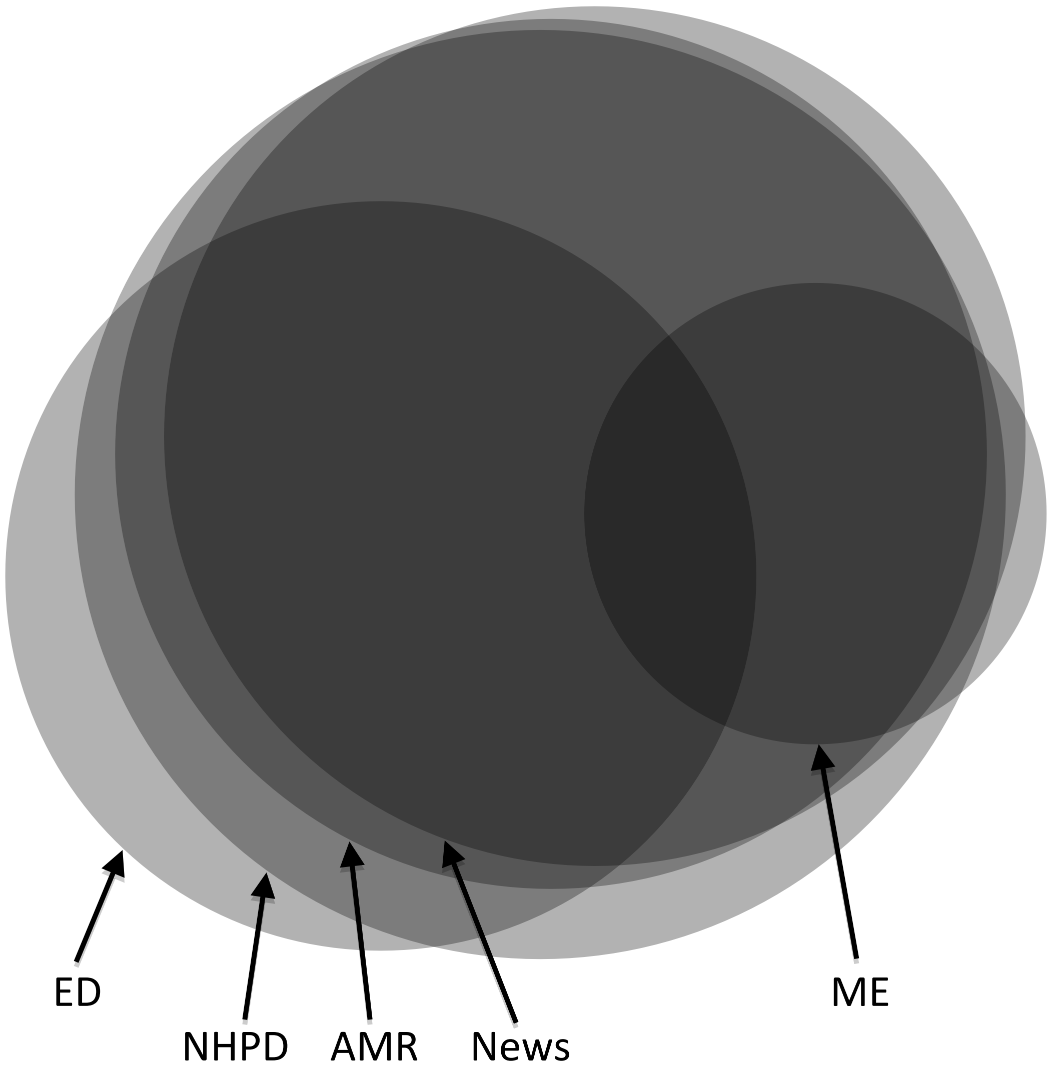
[2]These are victims that are found only in a single data set. For example, the table shows that 4 victims in the NHPD data were not found in any of the other data sets.

[3-4]These are the results from the ED with the 3 equivocal cases excluded or included, respectively.

Log-Linear Capture-Recapture Models

As illustrated above, there was considerable lack of overlap between the various data sources. Only the emergency department data captured injuries that did not show up in any other source. However, there were still a variety of different capture profiles. The overlap between sources can best be viewed as an area-proportional Venn diagram (Fig. 1). This figure demonstrates that there is substantial redundancy between the NHPD, AMR, and the news media. The medical examiner and the emergency department, however, capture slightly different segments of the injured population.

Figure 1: Proportional Venn diagram showing overlap of the five data sources.



The overlap shown in this diagram was then modeled using log-linear methods. As mentioned above, there were several ED records that we were neither able to completely exclude from our analysis or determine that they met our inclusion criteria. Therefore, we analyzed the data both including and excluding these records.

Complete Data Set

The complete data set included 49 gunshot injuries. Exploratory graphing for heterogeneity (Fig. 2) showed a non-linear form, indicating heterogeneity between capture probabilities.(129) The closed log-linear models were then evaluated by AIC, BIC, and residual boxplots. AIC and boxplots both indicated that the model "Mth Chao" fit best (see Table 2 and Fig. 3). This is one of several models available in the *Rcapture* package. The BIC indicated a marginally better fit for the model "Mt." However, we selected "Mth Chao" as the best model because—in addition to having better AIC and residuals—it allowed for heterogeneity among capture occasions as well as among individuals captured. This model estimated abundance of gunshot injuries at 49.7. The 95% confidence interval ranged from 49 to 52.3. Note that the inferior confidence interval was fixed to the actual number of recorded gunshot injuries. These results are shown below in Table 4 where they are compared to the abridged data set.

Figure 2: Exploratory Heterogeneity Graph of the complete data set.

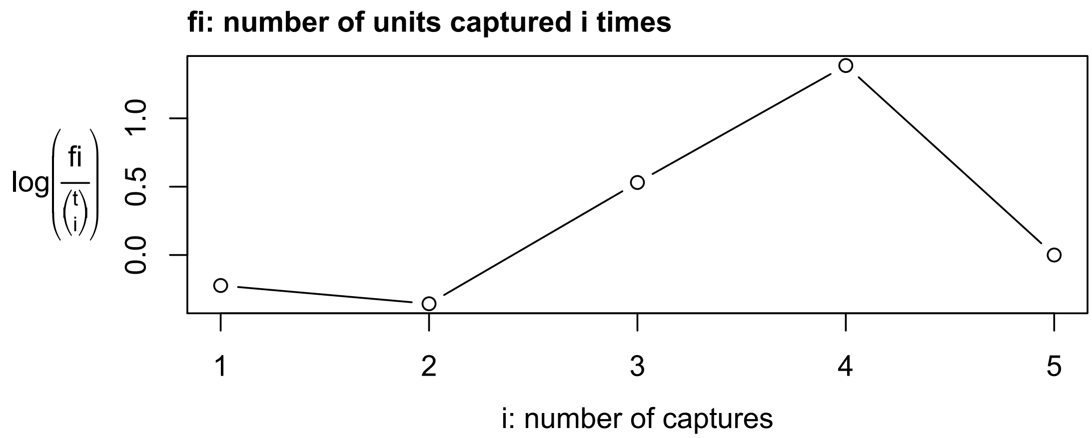
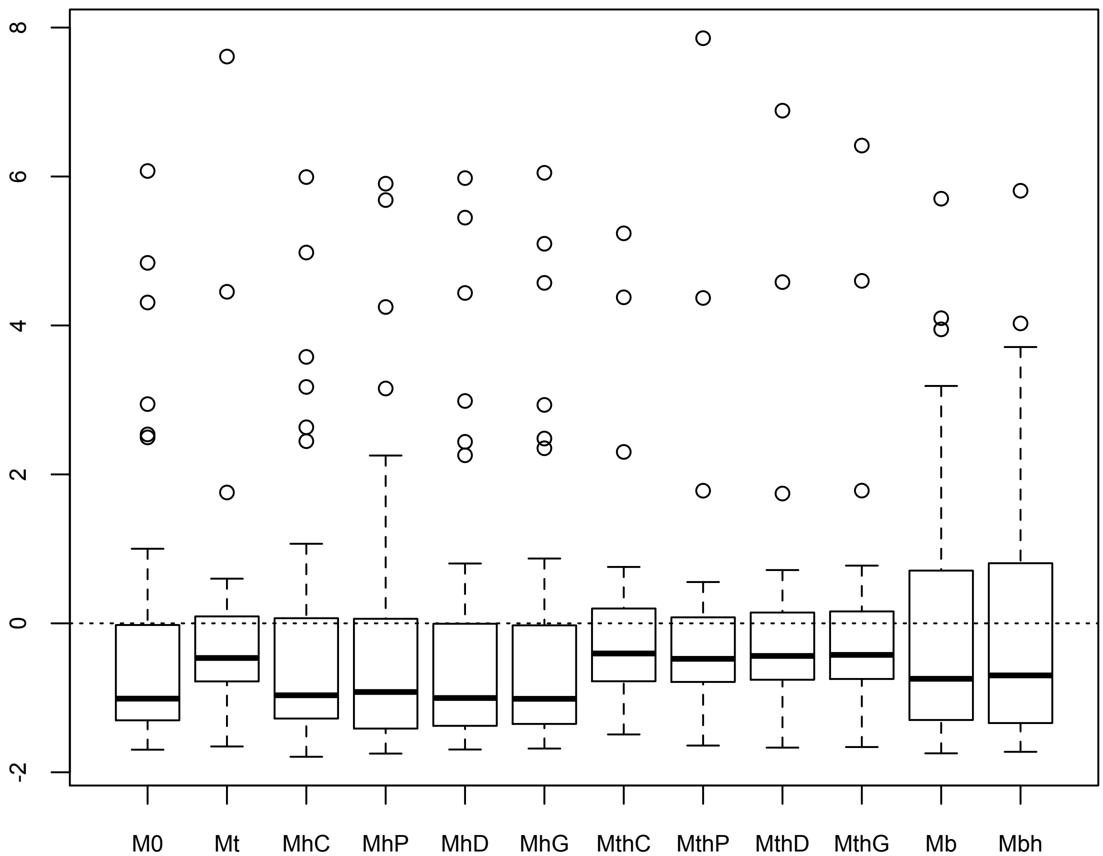


TABLE 2. MODEL RESULTS: CLOSED LOG-LINEAR MODELS FOR THE COMPLETE DATA

MODEL	ABUNDANCE	STDERR	DEVIANCE	DF	AIC	BIC
M0	49.37	0.63	113.21	29	152.38	156.16
Mt	49.14	0.39	53.33	25	100.49	111.84
Mh Chao (LB)	49.69	1.09	112.83	28	154.00	159.67
Mh Poisson2	49.14	0.39	110.71	28	151.87	157.55
Mh Darroch	49.15	0.42	112.34	28	153.51	159.18
Mh Gamma3.5	49.17	0.47	112.83	28	154.00	159.67
Mth Chao (LB)	49.70	1.11	51.03	24	100.19	113.44
Mth Poisson2	49.13	0.37	53.29	24	102.46	115.70
Mth Darroch	49.23	0.55	53.10	24	102.27	115.51
Mth Gamma3.5	49.41	0.86	52.76	24	101.92	115.16
Mb	49.05	0.22	107.64	28	148.80	154.48
Mbh	49.02	4.59	107.29	27	150.45	158.02

Figure 3: Boxplots of Pearson residuals of the models fit by *Rcapture* for the complete data set.



Abridge Data Set

The abridged data set was the same as the complete data set, but with the 3 equivocal injuries removed, giving 46 captured gunshot injuries. Exploratory graphing for heterogeneity showed a non-linear form,(Fig. 4) again indicating heterogeneity between capture probabilities. The AIC, BIC, and residual boxplot all indicated that model "Mt" and "Mth Chao" were equivalent as the best models (see

Table 3 and Fig. 5). Therefore, in order to have consistency with the above model of the complete data and to allow for heterogeneity among capture occasions as well as among individuals captured, we again chose "Mth Chao" as the best model. This model estimated abundance of gunshot injuries at 46.1. The 95% confidence interval ranged from 46 to 46.9. Note again that the inferior confidence interval was fixed to the actual number of recorded gunshot injuries. These results are shown below in Table 4 where they are compared to the complete data set.

Figure 4: Exploratory Heterogeneity Graph of the abridged data set.

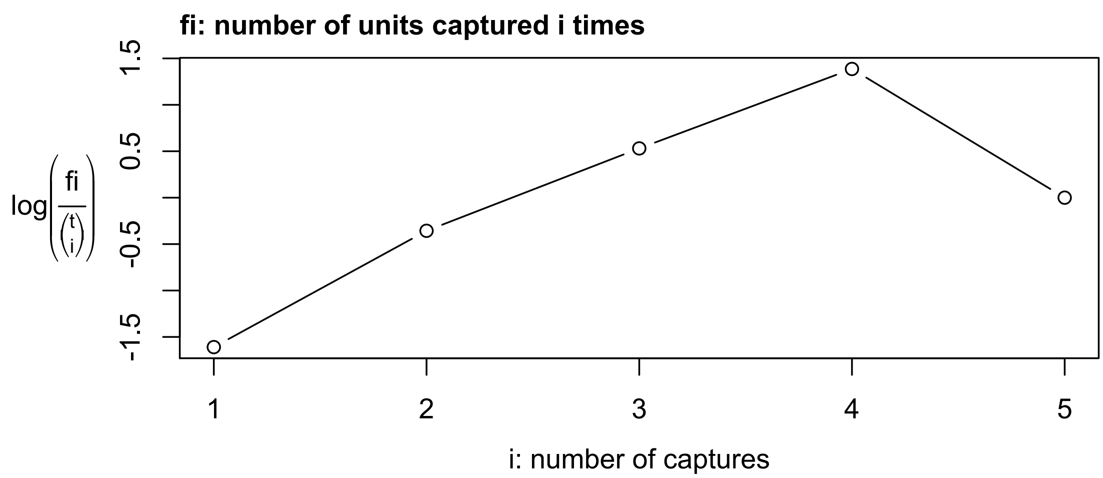


TABLE 3. MODEL RESULTS: CLOSED LOG-LINEAR MODELS FOR THE ABRIDGED DATA

MODEL	**ABUNDANCE**	**STDERR**	**DEVIANCE**	**DF**	**AIC**	**BIC**
M0	46.23	0.49	104.22	29	142.12	145.77
Mt	46.05	0.23	36.66	25	82.56	93.53
Mh Chao (LB)	46.23	0.49	104.22	29	142.12	145.77
Mh Poisson2	46.03	0.17	97.14	28	137.04	142.52
Mh Darroch	46.01	0.10	98.78	28	138.67	144.16
Mh Gamma3.5	46.00	0.07	99.56	28	139.46	144.95
Mth Chao (LB)	46.05	0.23	36.66	25	82.56	93.53
Mth Poisson2	46.02	0.14	35.22	24	83.12	95.92
Mth Darroch	46.01	0.12	35.92	24	83.82	96.62
Mth Gamma3.5	46.01	0.12	36.16	24	84.06	96.86
Mb	46.02	0.16	99.00	28	138.90	144.39
Mbh	46.03	3.88	98.98	27	140.88	148.20

Figure 5: Boxplots of Pearson residuals of the models fit by *Rcapture* for the abridged data set.

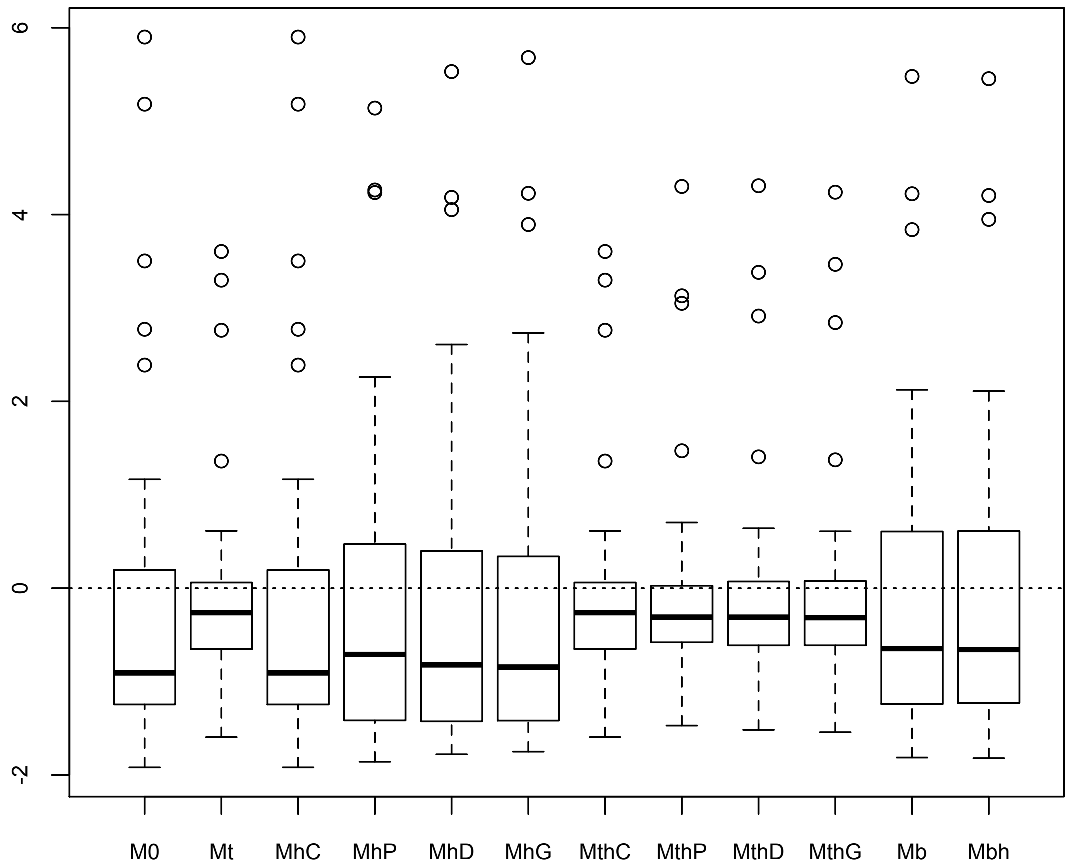


TABLE 4. SUMMARY OF ANALYSES: SELECTED MODEL RESULTS SHOWING ESTIMATED INJURIES

	RECORDED INJURIES	**ESTIMATED INJURIES**	**STANDARD ERROR**	**95% CI**
Complete Data Set	49	49.7	1.1	49 – 52.3
Abridged Data Set	46	46.1	0.2	46 – 46.9

Discussion

In this study we have demonstrated the application of a novel method to ascertain gunshot injuries. By combining data from 5 distinct sources we have generated a more complete picture than could be had from any single source alone. In addition, we have applied log-linear modeling techniques to estimate the number of additional uncounted gunshot injuries that occurred during our study period by operationalizing the overlap instead of simply matching, un-duplicating, and adding the data sources together.

The most complete single source was the records of the New Haven Police Department, recording 43 incidents. However, similar to Kellerman et al,(21) who found that 9% of cases lacked corresponding police reports, 6 out of our 49 cases (12%) were not included in police records. There are a variety of possible explanations for this. At the most basic level, there may have been clerical errors that led to those records being misplaced or mis-categorized. There is also a possibility that some were *intentionally* missing from police records because the police may have determined that the incident did not meet their definition of a gunshot injury. In fact, for several incidents that did not have a police record, other sources mentioned that the police were on scene. This implies that the police did know about some of these incidents, but for some reason a report was not generated or did not make it into our data set. In the case of emergency department records, it is also possible that some cases were not known to the police because ED providers may have neglected to provide notification.

In addition, it was not always clear if an incident represented a gunshot injury. In at least one case, a victim reported having been shot, but police and medical personnel determined that it was unlikely the victim had been shot. In several other cases, victims claimed that they had *not* been shot, but police and medical personnel felt that their wounds were most likely due to gunshots. In our analyses we accepted the classification of wounds as done by police and medical personnel. However, it should be kept in mind that this sometimes conflicted with victim reports and may not have always been completely accurate. This difficulty in determining with 100% certainty if a wound is from a firearm may explain some of the inconsistencies between data sources.

There was also a wide range of quality in the data from different sources. The medical examiner data, for example, was consistently of the highest quality for the purposes of our study. Names and birthdates appeared to be correct. There were extensive details about the location, date, time, and mechanism of the injury. There was a similar wealth of detail regarding the death following the injury. The main drawback, however, of the medical examiner data was, of course, that it is limited to fatal cases. The news sources seemed to mostly reflect police reports. In fact, most news sources directly attributed their information to the New Haven Police

Department. Regardless, news sources served to provide additional details regarding the specific details of incidents.

The data from the emergency department presented a large number of problems. Electronic medical records contain a wealth of information, but the other side of that coin is that things can be encoded in a variety of ways, making it difficult to search the medical record for events that are not coded in a uniform way. Therefore, our query may not have captured all of the gunshot wounds that presented to the emergency department. Similarly, there are many clerical errors in the emergency department data. For example, one case was known to involve a BB gun rather than a firearm based on data from the NHPD and AMR; the emergency department data coded this incident as a "gunshot wound."

Very few fatal gunshot wounds showed up in the emergency department data. Only 1 such wound was present in the emergency department data, despite the fact that every other data set recorded 8 or 9 fatal gunshot wounds. This occurred despite the fact that news media, AMR, and medical examiner records mention these patients being transported to the YNHH emergency department for treatment before they died. Perhaps this occurred because fatal gunshot wounds were more likely to be coded in the EMR as "cardiac arrest" or "hemorrhagic shock" than as "gunshot wound." It may also reflect the fact that fatal cases in the emergency department are likely to involve full trauma responses, which draw resources from across the emergency department, thus interfering with the work flow in the rest of the department and perhaps increasing the probability of clerical errors.

In addition to the limitations of our data sources described above, our research has some methodological limitations. For purposes of analysis, we have used a narrow case definition. Gun events have the ability to create an impact without discharge. For example, a victim may be threatened or coerced by an individual with a gun even without the gun being fired. Similarly, a gun may be discharged and not cause an injury or cause an injury that does not lead to the victim presenting for medical care. These are all examples of events that our study design would not be able to detect. Thus our estimates are conservative and exclude incidents involving non-firearm guns (e.g. airguns) and events in which a gun was used for intimidation but without actual discharge. A broader definition of gun-related incidents would introduce its own complexities into the study. Injuries due to non-firearm guns may not be consistently coded the same way in medical documentation. They also may not always be classified as “shootings” by police records. Similarly, events that involve mental trauma rather than physical injury may be less likely to come to medical or legal attention and therefore would be more difficult to measure, requiring different methods than those we have applied here.

Despite the presumption that capture-recapture might overestimate cases, our log-linear models demonstrate that there were very few uncounted cases. This may in fact reflect near complete case ascertainment by the combined data sets. It also could be a function of "structural zeroes"(23) in the population we are trying to capture. Structural zeroes refer to cases that are missed because they inherently *cannot* be captured by a given data set. They are contrasted with "random zeroes" that are missed due to chance. In other words, each victim should have a positive probability of being captured by each data set. If certain victims are *systematically* missed then they cannot be accounted for by the model.(23) Another way to view this concept, is to realize that log-linear modeling depends on the assumption that the uncounted cases are not qualitatively different from the counted cases. For example, above we described the hypothetical case of a victim who decided to hide his injuries because he feared legal or economic consequences. Unless this sort of victim has a non-zero probability of being counted by our data sources, the capture-recapture method cannot account for him.

Attempting to estimate these "structural zeroes" may be difficult. Collecting additional, independent data sets may provide more accurate point estimates. We initially hoped to collect data from social media sources, with the hopes that some individuals who did not present for medical care may have mentioned their injuries on social media platforms. An initial exploration of social media (not included in the above study) did reveal several mentions of gunshot injuries.

In summary, our study has demonstrated a method of estimating gunshot injury prevalence using multiple sources and capture-recapture methods. Withstanding the limitations, capture-recapture is superior to existing methods. The matched data set including results from five sources demonstrated that no single source captured all firearm injuries during the study period. Furthermore, the combined data set highlighted deficiencies in individual sources and provided a richer picture of each case. Interventions aimed at reducing the public health impact of gunshot injuries should be based on data using multiple sources and capture-recapture methods to maximize accuracy of prevalence estimates.

References

1. Cook PJ, and Ludwig J. *Guns in America: national survey on private ownership and use of firearms.* US Department of Justice, Office of Justice Programs, National Institute of Justice; 1997.
2. Wright JD, and Rossi PH. Washington, DC: United States Department of Justice, National Institute of Justice; 1985.
3. Kleck G, and Gertz M. Armed resistance to crime: the prevalence and nature of self-defense with a gun. *J Crim L & Criminology.* 1995;86(150.
4. McDowall D, and Wiersema B. The incidence of defensive firearm use by US crime victims, 1987 through 1990. *American Journal of Public Health.* 1994;84(12):1982-4.
5. Hepburn L, Miller M, Azrael D, and Hemenway D. The US gun stock: results from the 2004 national firearms survey. *Injury Prevention.* 2007;13(1):15-9.
6. Richardson EG, and Hemenway D. Homicide, suicide, and unintentional firearm fatality: comparing the United States with other high-income countries, 2003. *The Journal of Trauma and Acute Care Surgery.* 2011;70(1):238-43.
7. Goss KA. *Disarmed: The missing movement for gun control in America.* Princeton University Press; 2010.
8. Finkelhor D. *Children's exposure to violence: A comprehensive national survey.* DIANE Publishing; 2009.
9. Finkelhor D, Ormrod R, Turner H, and Hamby SL. The victimization of children and youth: A comprehensive, national survey. *Child maltreatment.* 2005;10(1):5-25.
10. Finkelhor D, Turner H, Ormrod R, and Hamby SL. Violence, abuse, and crime exposure in a national sample of children and youth. *Pediatrics.* 2009;124(5):1411-23.
11. Nance ML, Carr BG, Kallan MJ, Branas CC, and Wiebe DJ. Variation in pediatric and adolescent firearm mortality rates in rural and urban US counties. *Pediatrics.* 2010;125(6):1112-8.
12. Nance ML, Denysenko L, Durbin DR, Branas CC, Stafford PW, and Schwab CW. The rural-urban continuum: variability in statewide serious firearm injuries in children and adolescents. *Archives of pediatrics & adolescent medicine.* 2002;156(8):781-5.
13. Cook PJ, and Ludwig J. The social costs of gun ownership. *Journal of Public Economics.* 2006;90(1):379-91.
14. Cook PJ, and Ludwig J. The costs of gun violence against children. *The Future of Children.* 2002:87-99.
15. Ludwig J, and Cook PJ. *Evaluating gun policy: Effects on crime and violence.* Brookings Institution Press; 2003.
16. Ludwig J, and Cook PJ. The benefits of reducing gun violence: Evidence from contingent-valuation survey data. *Journal of risk and Uncertainty.* 2001;22(3):207-26.
17. Wadman M. Firearms research: The gun fighter. *Nature.* 2013;496(7446):412-5.

18. Kellermann AL, Bartolomeos K, Fuqua-Whitley D, Sampson TR, and Parramore CS. Community-level firearm injury surveillance: local data for local action. *Annals of emergency medicine.* 2001;38(4):423-9.
19. Pyramyd Air. http://www.pyramydair.com/. Accessed February 3, 2015.
20. Conn. Gen. Stat. § 19a-490f. 2013.
21. Kellermann AL, Rivara FP, Lee RK, Banton JG, Cummings P, Hackman BB, and Somes G. Injuries due to firearms in three cities. *New England Journal of Medicine.* 1996;335(19):1438-44.
22. HBO, The Wire. "The Cost" plot synopsis. http://www.hbo.com/ - /the-wire/episodes/1/10-the-cost/synopsis.html. Accessed December 22, 2014.
23. Chao A, Tsay PK, Lin SH, Shau WY, and Chao DY. The applications of capture-recapture models to epidemiological data. *Statistics in medicine.* 2001;20(20):3123-57.
24. McCarty DJ, Tull ES, Moy CS, Kwoh CK, and LaPorte RE. Ascertainment corrected rates: applications of capture-recapture methods. *International Journal of Epidemiology.* 1993;22(3):559-65.
25. Chang YF, McMahon JE, Hennon DL, LaPorte RE, and Coben JH. Dog bite incidence in the city of Pittsburgh: a capture-recapture approach. *Am J Public Health.* 1997;87(10):1703-5.
26. Roberts I, and Scragg R. Application of capture-recapture methodology to estimate the completeness of child injury surveillance. *Journal of paediatrics and child health.* 1994;30(6):513-4.
27. Post LA, Mezey NJ, Maxwell CD, and Rhodes K. Using capture-recapture to estimate the prevalence of intimate partner violence: The gender symmetry debate. *International Journal of Science in Society.* 2011;2(2).
28. Amstrup SC, McDonald TL, and Manly BF. *Handbook of capture-recapture analysis.* Princeton University Press; 2010.
29. National Research Council. *Firearms and Violence: A Critical Review.* Washington, DC: The National Academies Press; 2005.
30. Weil DS, and Hemenway D. Loaded guns in the home: analysis of a national random survey of gun owners. *JAMA.* 1992;267(22):3033-7.
31. Miller M, and Hemenway D. Guns and suicide in the United States. *New England Journal of Medicine.* 2008;359(10):989-91.
32. Hemenway D. *Private guns, public health.* University of Michigan Press; 2004.
33. Hemenway D. Private Guns. *Public Health.* 2004:367-79.
34. Miller M, Hemenway D, and Wechsler H. Guns and gun threats at college. *Journal of American College Health.* 2002;51(2):57-65.
35. Bergstein JM, Hemenway D, Kennedy B, Quaday S, and Ander R. Guns in young hands: a survey of urban teenagers' attitudes and behaviors related to handgun violence. *Journal of Trauma and Acute Care Surgery.* 1996;41(5):794-8.
36. Hemenway D. Survey research and self-defense gun use: an explanation of extreme overestimates. *J Crim L & Criminology.* 1996;87(1430.
37. Hemenway D, Vriniotis M, and Miller M. Is an armed society a polite society? Guns and road rage. *Accident Analysis & Prevention.* 2006;38(4):687-95.
38. Hemenway D, Solnick SJ, and Azrael DR. Firearm training and storage. *Jama.* 1995;273(1):46-50.

39. Rothman EF, Hemenway D, Miller M, and Azrael D. Batterers' use of guns to threaten intimate partners. *Journal of the American Medical Women's Association (1972).* 2004;60(1):62-8.
40. Miller M, Azrael D, and Hemenway D. Rates of household firearm ownership and homicide across US regions and states, 1988-1997. *American Journal of Public Health.* 2002;92(12):1988-93.
41. Young JT, Hemenway D, Blendon RJ, and Benson JM. Trends: guns. *Public Opinion Quarterly.* 1996:634-49.
42. Blendon RJ, Young JT, and Hemenway D. The American public and the gun control debate. *JAMA.* 1996;275(22):1719-22.
43. Hemenway D, and Miller M. Firearm availability and homicide rates across 26 high-income countries. *Journal of Trauma and Acute Care Surgery.* 2000;49(6):985-8.
44. Miller M, Lippmann SJ, Azrael D, and Hemenway D. Household firearm ownership and rates of suicide across the 50 United States. *Journal of Trauma and Acute Care Surgery.* 2007;62(4):1029-35.
45. Miller M, Azrael D, and Hemenway D. Firearm availability and unintentional firearm deaths, suicide, and homicide among 5–14 year olds. *Journal of Trauma and Acute Care Surgery.* 2002;52(2):267-75.
46. Hemenway D, Prothrow-Stith D, Bergstein JM, Ander R, and Kennedy BP. Gun carrying among adolescents. *Law and Contemporary Problems.* 1996:39-53.
47. Hemenway D, Solnick SJ, and Azrael DR. Firearms and community feelings of safety. *J Crim L & Criminology.* 1995;86(121.
48. Hemenway D, Shinoda-Tagawa T, and Miller M. Firearm availability and female homicide victimization rates among 25 populous high-income countries. *Journal of the American Medical Womens Association.* 2002.
49. Duggan M. National Bureau of Economic Research; 2000.
50. Kellerman A, and Reay D. Gun Ownership Is Not an Effective Means of Self-Defense (From Gun Control, P 171-175, 1992, Charles P Cozic, ed.--See NCJ-160164). 1992.
51. Rivara FP, and Koepsell TD. Money for Guns: Evaluation of the Seattle Gun Buy-Back Program Charles M Callahan, University of Rochester. *Under fire: gun buy-backs, exchanges and amnesty programs.* 1996:85.
52. Martinez Jr R. *Firearms and Violence: Oxford Bibliographies Online Research Guide.* Oxford University Press; 2010.
53. Loomis D, Marshall SW, and Ta ML. Employer policies toward guns and the risk of homicide in the workplace. *American journal of public health.* 2005;95(5):830.
54. Miller M, Azrael D, and Hemenway D. Household firearm ownership and suicide rates in the United States. *Epidemiology.* 2002;13(5):517-24.
55. Miller M, Azrael D, and Hemenway D. Firearm availability and suicide, homicide, and unintentional firearm deaths among women. *Journal of Urban Health.* 2002;79(1):26-38.
56. Hepburn LM, and Hemenway D. Firearm availability and homicide: A review of the literature. *Aggression and Violent Behavior.* 2004;9(4):417-40.

57. Richardson EG, and Hemenway D. Homicide, suicide, and unintentional firearm fatality: comparing the United States with other high-income countries, 2003. *Journal of Trauma and Acute Care Surgery.* 2011;70(1):238-43.
58. Azrael D, and Hemenway D. 'In the Safety of Your Own Home': Results from a National Survey on Gun Use at Home. *Social Science & Medicine.* 2000;50(2):285-91.
59. Miller M, Hemenway D, and Azrael D. State-level homicide victimization rates in the US in relation to survey measures of household firearm ownership, 2001–2003. *Social science & medicine.* 2007;64(3):656-64.
60. Kellerman AL. *Understanding and preventing violence: a public health perspective.* National Criminal Justice Reference Service; 1994.
61. Olson L, Huyler F, Lynch AW, Fullerton L, Werenko D, Sklar D, and Zumwalt R. Guns, alcohol, and intimate partner violence: the epidemiology of female suicide in New Mexico. *Crisis: The Journal of Crisis Intervention and Suicide Prevention.* 1999;20(3):121.
62. Kellermann AL, and Rivara FP. Silencing the Science on Gun ResearchSilencing the Science on Gun Research. *JAMA.* 2013;309(6):549-50.
63. Jamieson C. Gun violence research: History of the federal funding freeze. *Psychological Science Agenda.* 2013;27(2).
64. Kopel D. How Your Tax Dollars Demonize Your Guns. http://www.nrapublications.org/index.php/9485/how-your-tax-dollars-demonize-your-guns/. Accessed August 1, 2014, 2014.
65. Braman D, and Kahan DM. Overcoming the Fear of Guns, the Fear of Gun Control, and the Fear of Cultural Politics: Constructing a Better Gun Debate. *Emory Law Journal.* 2006;55(2012-58.
66. Cook PJ, Ludwig J, and Hemenway D. The gun debate's new mythical number: How many defensive uses per year? *Journal of Policy Analysis and Management.* 1997;16(3):463-9.
67. Braman D, Kahan DM, and Grimmelmann J. Modeling facts, culture, and cognition in the gun debate. *Social Justice Research.* 2005;18(3):283-304.
68. Frosch D, and Johnson K. Gunman kills 12 in Colorado, reviving gun debate. *The New York Times.* 2012.
69. Friedman RA. In gun debate, a misguided focus on mental illness. *The New York Times.* 2012.
70. Frosch D, and Johnson K. July 20, 2012."Gunman kills 12 in Colorado, reviving gun debate.". *The New York Times Retrieved on January.* 2014;14(
71. Blocher J. Gun Rights Talk. *BUL Rev.* 2014;94(813.
72. Martin DS, Murray D, O'Neill MA, MacCarthy M, and Gogue J. Target shooting as a serious leisure pursuit–an exploratory study of the motivations driving participant engagement. *World Leisure Journal.* 2014;56(3):204-19.
73. Brasted M. The NRA and Polarization in the Gun Debate. *The Socjournal.* 2014.
74. Branas CC, Richmond TS, Culhane DP, Ten Have TR, and Wiebe DJ. Investigating the link between gun possession and gun assault. *Am J Public Health.* 2009;99(11):2034-40.

75. Consolidated Appropriations Act 2012, Pub L No. 112-74. http://www.gpo.gov/fdsys/pkg/PLAW-112publ74/pdf/PLAW-112publ74.pdf. December 2011. Accessed December 24, 2014.
76. Office of the Press Secretary, The White House. New Executive Actions to Reduce Gun Violence. August 29, 2013. http://www.whitehouse.gov/the-press-office/2013/08/29/fact-sheet-new-executive-actions-reduce-gun-violence. Accessed 12/10/2014.
77. Ohlheiser A. *Slate*. 2013.
78. Senate, Office of Edward J. Markey. "Sen. Markey & Rep. Maloney Introduce Legislation to Fund Gun Violence Research." May 21, 2014. http://www.markey.senate.gov/news/press-releases/sen-markey-and-rep-maloney-introduce-legislation-to-fund-gun-violence-research. Accessed December 15, 2014.
79. Raile ED, Raile AN, Salmon CT, and Post LA. Defining Public Will. *Politics & Policy.* 2014;42(1):103-30.
80. Post LA, Raile AN, and Raile ED. Defining political will. *Politics & Policy.* 2010;38(4):653-76.
81. Post LA, Salmon CT, and Raile AN. Using public will to secure political will. *REFORM.* 2008:113.
82. Chandran A, Puvanachandra P, and Hyder AA. Commentary: Prevention of violence against children: A framework for progress in low-and middle-income countries. *Journal of public health policy.* 2010;32(1):121-34.
83. Farooq U, Majeed M, Bhatti JA, Khan JS, Razzak JA, and Khan MM. Differences in reporting of violence and deliberate self harm related injuries to health and police authorities, Rawalpindi, Pakistan. *PloS one.* 2010;5(2):e9373.
84. Gerstenberger CB, and Williams KR. Gender and intimate partner violence: does dual arrest reveal gender symmetry or asymmetry? *Journal of interpersonal violence.* 2012.
85. Guberek T, Guzmán D, Price M, Lum K, and Ball P. To count the uncounted: An estimation of lethal violence in Casanare. *A Report by the Benetech Human Rights Program.* 2010.
86. Oosterlee A, Vink RM, and Smit F. Prevalence of family violence in adults and children: estimates using the capture–recapture method. *The European Journal of Public Health.* 2009;19(6):586-91.
87. Silva R, and Price M. Trade-offs in using indirect sampling to measure conflict violence. *JAMA: the journal of the American Medical Association.* 2011;306(5):547.
88. Manrique-Vallier D, Price ME, and Gohdes A. Multiple Systems Estimation Techniques for Estimating Casualties in Armed Conflicts. *Counting Civilian Casualties: An Introduction to Recording and Estimating Nonmilitary Deaths in Conflict.* 2013:165.
89. Rossi G, Pepe P, Curzio O, and Marchi M. Generalized linear models and capture-recapture method in a closed population: strengths and weaknesses. *Statistica.* 2010;70(3):371-90.

90. Goudie IBJ, and Goudie M. Who Captures the Marks for the Petersen Estimator? *Journal of the Royal Statistical Society Series A (Statistics in Society).* 2007;170(3):825-39.
91. Chao A, and Huggins RM. In: Amstrup SC, McDonald TL, and Manly BF eds. *Handbook of Capture-Recapture Analysis*. Princeton University Press; 2005.
92. Cormack RM. Log-linear models for capture-recapture. *Biometrics.* 1989:395-413.
93. Fienberg SE. The multiple recapture census for closed populations and incomplete 2k contingency tables. *Biometrika.* 1972;59(3):591-603.
94. Brown J, Abbott O, and Diamond I. Dependence in the 2001 One-Number Census Project. *Journal of the Royal Statistical Society Series A (Statistics in Society).* 2006;169(4):883-902.
95. Cormack R. The statistics of capture-recapture methods. *Oceanogr Mar Biol Ann Rev.* 1968;6(455-506.
96. Capture-recapture and multiple-record systems estimation I: History and theoretical development. International Working Group for Disease Monitoring and Forecasting. *American journal of epidemiology.* 1995;142(10):1047-58.
97. Sekar CC, and Deming WE. On a method of estimating birth and death rates and the extent of registration. *Journal of the American Statistical Association.* 1949;44(245):101-15.
98. Anderson M, and Fienberg SE. *Who counts?: The politics of census-taking in contemporary America.* Russell Sage Foundation; 1999.
99. Brunell TL. Using statistical sampling to estimate the US population: the methodological and political debate over census 2000. *PS: Political Science & Politics.* 2000;33(04):775-82.
100. Hannah MG. Sampling and the politics of representation in US Census 2000. *ENVIRONMENT AND PLANNING D.* 2001;19(5):515-34.
101. Brown LD, Eaton ML, Freedman DA, Klein SP, Olshen RA, Wachter KW, Wells MT, and Ylvisaker D. Statistical controversies in Census 2000. *Jurimetrics.* 1999:347-75.
102. Wittes J, and Sidel VW. A generalization of the simple capture-recapture model with applications to epidemiological research. *Journal of chronic diseases.* 1968;21(5):287-301.
103. International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation II: Applications in human diseases. *American journal of epidemiology.* 1995;142(1059–68.
104. Lewis CE, and Hassanein KM. The relative effectiveness of different approaches to the surveillance of infections among hospitalized patients. *Medical care.* 1969;7(5):379-84.
105. Morrison A, and Stone DH. Capture-recapture: a useful methodological tool for counting traffic related injuries? *Injury prevention : journal of the International Society for Child and Adolescent Injury Prevention.* 2000;6(4):299-304.
106. LaPorte RE, Dearwater SR, Chang YF, Songer TJ, Aaron DJ, Anderson RL, and Olsen T. Efficiency and accuracy of disease monitoring systems: application of capture-recapture methods to injury monitoring. *American journal of epidemiology.* 1995;142(10):1069-77.

107. Tercero F, and Andersson R. Measuring transport injuries in a developing country: an application of the capture-recapture method. *Accident; analysis and prevention.* 2004;36(1):13-20.
108. Boden LI, and Ozonoff A. Capture-recapture estimates of nonfatal workplace injuries and illnesses. *Annals of epidemiology.* 2008;18(6):500-6.
109. Rosenman KD, Kalush A, Reilly MJ, Gardiner JC, Reeves M, and Luo Z. How much work-related injury and illness is missed by the current national surveillance system? *Journal of occupational and environmental medicine / American College of Occupational and Environmental Medicine.* 2006;48(4):357-65.
110. Jarvis SN, Lowe PJ, Avery A, Levene S, and Cormack RM. Children are not goldfish--mark/recapture techniques and their application to injury data. *Injury prevention : journal of the International Society for Child and Adolescent Injury Prevention.* 2000;6(1):46-50.
111. International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation I: History and theoretical development. *American journal of epidemiology.* 1995;142(1047–58.
112. Vaissade L, and Legleye S. Capture–recapture estimates of the local prevalence of problem drug use in six French cities. *The European Journal of Public Health.* 2009;19(1):32-7.
113. Lum K, Price M, Guberek T, and Ball P. 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.* 2010;1(1).
114. U.S. Census Bureau. "American Fact Finder". http://factfinder.census.gov/faces/nav/jsf/pages/community_facts.xhtml. Accessed December 14, 2014.
115. Garcia M. *Creating a Healthy and Safe City: The Impact of Violence in New Haven.* New Haven, CT: New Haven Public Health Department.; 2011.
116. Yale-New Haven Hospital. "Yale-New Haven Hospital History". http://www.ynhh.org/about-us/our-history.aspx. Accessed December 20, 2014.
117. Yale Emergency Medicine Residency. "Yale-New Haven Hospital". http://www.medicine.yale.edu/emergencymed/residency/hospital/ynhh.aspx. Accessed December 22, 2014.
118. New Haven Police Department. "DECEMBER 2012 and YEAR TO DATE - UNIFORM CRIME REPORTS - PART I CRIME STATISTICS". http://www.cityofnewhaven.com/Police/2012_Year_End_Statistics.asp. Accessed December 15, 2014.
119. American Medical Response. http://www.amr.net/Locations/Operations/Connecticut/New-Haven. Accessed December 22, 2014.
120. Office of the Chief Medical Examiner. http://www.ct.gov/ocme/site/default.asp. Accessed December 26, 2014.
121. New Haven Independent. http://www.newhavenindependent.org/. Accessed December 22, 2014.
122. New Haven Register. http://www.nhregister.com. Accessed December 26, 2014.

123. Hartford Courant. http://www.courant.com. Accessed December 28, 2014.
124. WTNH. http://wtnh.com/. Accessed December 26, 2014.
125. Fox CT. http://www.foxct.com. Accessed December 26, 2014.
126. Biroscak BJ, Smith PK, and Post LA. A practical approach to public health surveillance of violent deaths related to intimate partner relationships. *Public health reports (Washington, DC : 1974)*. 2006;121(4):393-9.
127. R Development Core Team. Vienna, Austria: R Foundation for Statistical Computing; 2014.
128. R Development Core Team. CRAN Task View: Analysis of Ecological and Environmental Data. http://cran.r-project.org/web/views/Environmetrics.html. Accessed December 21, 2014.
129. Baillargeon S, and Rivest L-P. Rcapture: loglinear models for capture-recapture in R. *Journal of Statistical Software.* 2007;19(5):1-31.
130. Burnham KP, and Anderson DR. Multimodel inference understanding AIC and BIC in model selection. *Sociological methods & research.* 2004;33(2):261-304.

Appendix A: Code for Statistical Analysis in R

```
################################################################
################################################################
### Capture Recapture Analysis of New Haven Firearms Injuries ###
################################################################


## Load the data as a 2 column table with headers, indicating individuals  and data sets
## that captured them
#deID <- read.csv("deidentified.abridged.csv" , header=TRUE, sep=",")
deID <- read.csv("deidentified.complete.csv" , header=TRUE, sep=",")
deID


## Load the fuzzySim package, in order to use splist2presabs to create presence-absence
## table
library(fuzzySim)
firearms <- splist2presabs(deID, sites.col="Match.ID", sp.col="Data.Set",
keep.n=FALSE)
firearms
## Remove Match.ID column:
firearms$Match.ID <- NULL
firearms


#### Load package Rcapture for capture-recapture analysis
```

```
library(Rcapture)

help(Rcapture)


## Explore the data

desc <- descriptive(firearms, dfreq=FALSE, dtype="hist")

plot(desc)

## (the ui plot is meaningless because it depends on the arbitrary ordering of the data

##sources)


## Fit the data as a closed population:

## Based on page 2 of the Baillargeon paper describing the Rcapture package, a priori,

##we can expect to have  a "th" model (heterogeneity and temporal effects). The b

##parameter (behavior) is meaningless because our captures have no defined temporal

##relationship

closed.firearms <- closedp(firearms)

closed.firearms

## export table

#write.table(closed.firearms$results, "Model.Results/Model.Results.txt", sep="\t")

## which model minimizes the AIC and/or BIC? That is the model with the best fit.


## evaluate the fit of the model by looking at plot of the residuals:

boxplot(closed.firearms)

## which model has the tightest resideuals?
```

```
## Select a model:
model.selected <- "Chao"


## Calculate the confidence intervals
CI.firearms <- closedpCI.t(firearms, m="Mth" , h=model.selected)
CI.firearms
plotCI(CI.firearms, main="Profile Likelihood Confidence Interal")


##################################################################
##### Create proportional Venn diagram ########################
##################################################################
library(venneuler)
gsw.venn <- venneuler(firearms)
gsw.venn$labels <- c("","","","","")
plot(gsw.venn, main= "", col="black", alpha=.3)
```