A QUANTITATIVE EVALUATION OF THREE MASS CASUALTY TRIAGE SYSTEMS IN COMMON USE IN THE UNITED STATES

A Master Thesis

Submitted to the Faculty

of

American Public University

by

Travis L. Houser

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Arts

April 2018

American Public University

Charles Town, WV

DEDICATION

I dedicate this thesis, firstly, to my wife and daughters. Their longsuffering patience, love, and encouragement were crucial to the completion of this work. Secondly, this thesis is dedicated to the men and women of the military medical services and civilian emergency response services. They begin each day with a quiet understanding that this may be the day that they need to go into harms way, so that others may live. It is my sincere hope that this work contributes in some small fashion to their ability to assist those in need.

ACKNOWLEDGEMENTS

I would like to thank the faculty members that have guided me through this process for their dedication and assistance. Dr. Randall Cuthbert, Dr. Michael Charter, and Dr. Randall Hanifen for guiding me through the initial stages of the program, Dr. Mehrdad Zarandi-Mahmoudi and Dr. Jessica Sapp for grounding me in statistics and public health, respectively, Dr. Edwin Bundy for preparing me for the rigors of the capstone process, and Dr. Erich Randall for helping me through the completion of the present work. I also could not have completed this research without the assistance of the Sailors assigned to Branch Health Clinic, Naval Station Norfolk, who volunteered their time to help complete this study.

ABSTRACT OF THE THESIS

A QUANTITATIVE EVALUATION OF THREE MASS CASUALTY TRIAGE SYSTEMS IN COMMON USE IN THE UNITED STATES

by

Travis L. Houser

American Public University System, January 15, 2018

Charles Town, WV

Professor Erich Randall, Thesis Professor

Since the development of the START triage algorithm 35 years ago, surprisingly little research has been performed comparing any of the various triage methods to an objective reference standard. Despite this lack of data, new triage methodologies continue to be developed to address perceived deficiencies in existing methods. A quantitative correlational study was conducted comparing the effectiveness of the START, SALT, and ESI triage systems, using the Injury Severity Score as a reference. While all three systems were found to exhibit some degree of correlation, the SALT algorithm was found to correlate most strongly to the Injury Severity Score, conformed to the expected distribution of patient acuity, and had the lowest rate of under- or over-triage of the three tested systems. Recommendations were provided for further academic study, modifications to the Model Uniform Core Criteria, and practical actions to be taken by the United States' emergency response apparatus. In the absence of further

study comparing SALT to other existing systems, a recommendation for widespread adoption of the SALT algorithm is withheld.

TABLE OF CONTENTS

A Quantitative Evaluation of Three Mass Casualty Triage Systems in Common Use in the United States

Introduction and Problem Summary

Background

In mass casualty incidents, decisions must be made about which patients will receive what treatment, and in which order they will do so, in order to assure that the greatest good is done for the greatest number. This is done by evaluating each patient's injuries and balancing their severity and risk of life-threat against the injuries of all of the other patients, as well as the resource constraints, both in terms of personnel and equipment, faced by those responding to the incident. This is accomplished through the use of any of a myriad of triage systems, each with their own unique algorithmic decision tree. These algorithms typically result in a four- or five-tiered categorization of patients, and these categories are utilized throughout the incident response to determine what interventions will be applied on-scene, the order in which patients should be transported to definitive care, and to what facility patients will be taken. These systems have their origins in the management of mass combat injuries on the battlefield, although the ultimate goal has been altered to meet the realities of the civilian environs.

What has translated over to the civilian world is the tiered classification of patients. Virtually all systems in use today utilize a four- or five-tiered classification system, often color coded to help ensure clarity in the middle of a chaotic situation. The only true variance among the various systems exists in two areas; whether the system utilizes an "Expectant" classification (meaning patients whose injuries are so severe that they are not expected to survive their injuries), and in the algorithmic decision process utilized to sort patients into the respective classifications.

At it's essence, triage is the means by which one makes decisions in a mass casualty incident that directly impact which patients will live and which ones will not, with the ultimate goal being to save as many lives as possible. In order to actually save the greatest number of patients, response personnel should employ a system that can be proven to be the most accurate at predicting patient mortality based on their injury, and it should do so using a method that can be replicated regardless of the clinical experience and knowledge of the person actually performing triage. This ensures that mass casualty incident management would function similarly to the Incident Command System, namely that it will provide a mechanism for a less experienced responder, even to the level of a layperson, can positively impact the incident until a more experienced responder can assume command.

Problem Summary

If one examines the idea of triage from the bottom up, from the perspective of the individuals whose lives are at stake in a mass casualty incident, it is shocking how little research has been attempted to determine which system is the most functionally accurate at predicting patient mortality. This seems an irresponsible oversight in light of the "big picture" question that underlies the entire concept of triage: How do we save the lives of the greatest number of patients? Typically, what research that has been conducted compares two or more triage methods to each other, but not to a defined reference standard. This is essentially the same as conducting an experiment with two experimental treatment protocols, but omitting the control; there is simply no means to tie the results of such a study to the specific variables that influenced those results.

As a result of this lack of effective research, it is not currently well understood to what degree the various triage systems correlate to the actual severity of the patient. This lack of a clear consensus on which system is empirically more effective at classifying patients into severity tiers has led to a proliferation of differing triage methodologies, beginning with the development of the START system in 1983 (US Department of Health and Human Services, 2017, September 29). As of this writing there are at least three triage systems in common usage in the United States: Simple Triage and Rapid Treatment (START), Sort, Assess, Lifesaving Interventions, Treatment/Transport (SALT), and the Emergency Severity Index (ESI). In addition to these three systems, there are at least three major protocols used internationally. Being the oldest, START is the best studied, and the most used, but even this system has seen very little in the way of quantitative evaluation of its effectiveness.

In the vacuum of empirical data, the use of any specific triage system by emergency medical services agencies or hospitals is dictated by policy or personal preference rather than by effectiveness. Crossing a jurisdictional boundary can mean transitioning to a completely different system. In the current mode of increasing reliance on interagency cooperation and national standardization of incident management protocols, this situation is potentially problematic (Streckbein, Kohlmann, Luxen, Birkholz, & Prückner, 2016), especially in major national or international incidents that draw responders from far-flung locales.

The aim of this study is to begin to fill this gap in the literature by evaluating three common triage systems used in the United States, and comparing them against the ‘gold standard’ of the Injury Severity Score (ISS), as calculated from the 2005 update of the Abbreviated Injury Scale (AIS(2005)). This well established (Bull, 1975) benchmark of injury severity has been shown repeatedly to correlate well to patient outcomes, but by its nature is unsuitable for use as a triage tool in and of itself. Establishing the effectiveness of these three systems relative to the ISS will enable medical directors and policy makers to base their decisions on empirically validated data, rather than precedent and intuitive preferences. This benchmark can also provide a launching point from which refinements to existing methods may be undertaken, as well as a comparison point for international or novel triage methodologies. If one of the tested systems can be shown clearly superior to the others, it can then be easily compared to other systems in use, both domestically and abroad, with the ultimate goal of arriving at an international consensus of what is the best method of triaging patients in a mass casualty incident. In an increasingly interconnected world, such a consensus would help facilitate global response to major disasters involving mass casualties.

Constraints

The systems that will be evaluated in this study do each have a relatively clear contextual setting in which they are most appropriately used. SALT, based off of its initial task of globally sorting patients by the ability to walk, and then by responsiveness, is uniquely well suited to evaluating patients at the point of injury. START is well suited to first *receiver* operations, such as at a casualty collection point, prior to delivering patients to definitive care. The ESI algorithm is expressly designed to be employed in an Emergency Room setting, and does not account for "expectant" or expired casualties, as these casualties would be filtered out of the treatment stream somewhere closer to the point of injury. Whether these qualitative differences have any bearing on the choice to employ one or another of the systems tested is necessarily outside the scope of this study. A second limitation of the study is that the ISS is fixed for each patient, and does not change over time without the advent of an additional injury. As the patient data utilized will be retrospective in nature, there is no way to account for the changes in patient mortality that occur over time, and thus evaluate whether the different systems correlate more or less effectively as the mass casualty incident unfolds.

In addition to these limitations, the work will deliberately avoid any distinction between penetrating and blunt trauma injuries. As the development of the Tactical Combat Casualty Care treatment algorithm has shown, there are significant differences in how the body responds to blunt versus penetrating mechanisms of injury, and the means employed to treat these opposing trauma models are very distinct. The possibility exists that there is a statistically significant difference in how these systems correlate with injury severity in the two trauma mechanisms that will not be evaluated in this work.

Lastly, this work will focus only on triage systems that are used in the United States. The reasoning for this is twofold: first, as the work is completed in the United States, the results of this study can potentially be put into practice more readily than in the international community, and second, the inclusion of triage systems used on the international stage expands the scope of the study enough to become unwieldy. Once a "gold standard" for triage is established from among the systems commonly used in the US, further work can evaluate that system against various methods used abroad, with the ultimate goal of arriving at an internationally agreed upon standard for conducting triage of mass casualty victims.

Literature Review

Need for the Study

There is a well-documented need for an evaluation of triage algorithm effectiveness, going back over a decade. Lerner, et al. (2008) noted that there was little consistency from jurisdiction to jurisdiction in which system was employed or how triage categories were defined, and little scientific data that could be used to validate any individual system of triage. Their study goes on to evaluate the available data and presents a consensus model that emerged from their multidisciplinary team, the SALT triage system. Similarly, Jenkins, et al. (2008) called for the development of a data-based triage method, extending their recommendation beyond only a national standard to that of an internationally accepted standard for conducting triage. Particularly telling, their study noted that, of six distinct triage systems evaluated, not one had been studied for reliability or were applicable to all hazards.

Most of the studies that have been conducted to date have not focused as directly on the effectiveness of the different triage algorithms themselves. One such study focused more on the physiologic markers used by four systems than on the overall efficacy of any of the systems themselves. Their conclusion was that, of the systems evaluated, the physiologic marker that most strongly correlated with severe injury was the Motor assessment of the Glasgow Coma Scale (GCS) (Garner, Lee, Harrison, & Schultz, 2001). Another study evaluated the correlation of triage classification as conducted by doctors, nurses, and paramedics, finding that doctors and nurses had a high level of agreement, but paramedics did not fare as well (Kilner, 2002). To date, there has only been one study conducted which attempted to compare triage systems to an objective reference standard (Silvestri, et al., 2017). This is, in part, due to the persistent lack of a true "gold standard" definition of each of the triage categories. This lack of an objective standard in the literature was recently corrected by Lerner, et al. (2015). Their work produced a consensus model suitable for retrospectively determining which category a given patient should have been placed into, although by its nature their work is not directly applicable to field usage, as many of their classification criteria rely on whether the patient needed specific interventions within given time intervals after arrival at definitive care. Lerner, et al.'s work (2015) was used by Silvestri, et al. as the reference standard for the comparison of the START and SALT algorithms (2017).

Proliferation of Triage Systems

In the absence of a definitive answer on the most accurate method of sorting patients during a mass casualty incident, many different systems have been proposed. START was originally developed in 1983 in Newport Beach, CA, and spawned a pediatric version in 1995, JumpSTART. These systems were modified in 1996 and 2001 respectively. The current state of the START algorithm is shown in Figure 1.

Garner, et. al. (2001) conducted the first comparative study involving the use of multiple systems, comparing START, the 1996 modification of the same, a British system known as Triage Sieve, and the Australian Careflight Triage system. Their study found that all tested systems produced similar results, but recommended Careflight Triage, as they believed it could be performed most rapidly in a field setting. Their study, as noted above, did not attempt to address the effectiveness of the tested triage systems at placing the patient into the correct category assignment, but rather focused on comparing how the various systems differed, and which of the examined physiologic markers were most predictive of patient acuity.

The Emergency Severity Index is a relatively new triage system, designed specifically for use in the Emergency Department (Wuerz, Milne, Eitel, Travers, & Gilboy, 2000). Of note, this system specifically addresses the number of resources required to treat the patient, something absent from previous models. Wuerz, Travers, Gilboy, Eitel, Rosenau, & Yazhari were able to show one year later that their system exhibited a strong degree of repeatability between different practitioners (2001), notably that the same patient had a strong likelihood of being triaged to the same category by personnel at two independent facilities. Currently on it's fourth revision and update, the current ESI algorithm is shown in Figure 2.

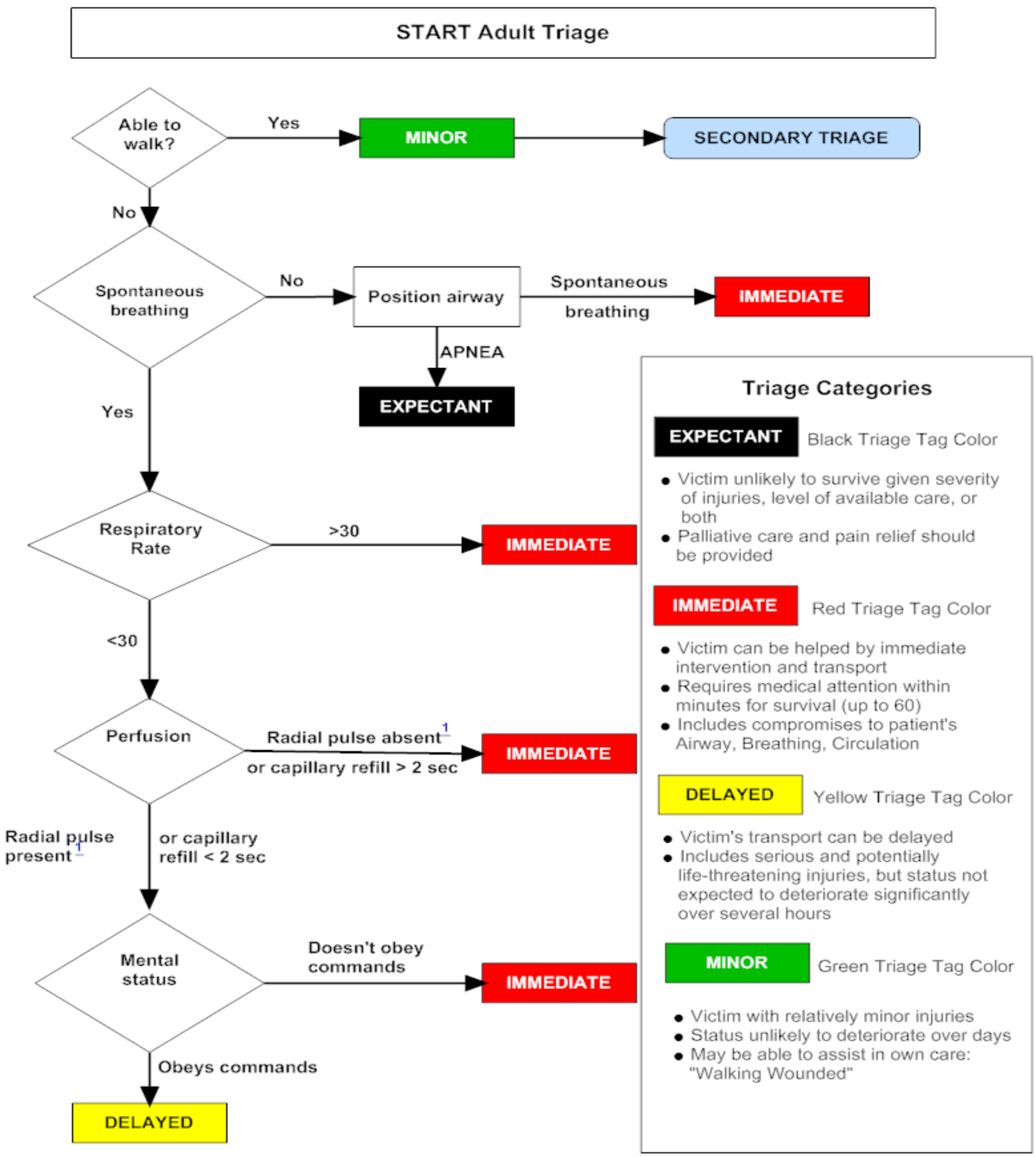

Figure 1: START Triage Algorithm. (START Adult Triage, 2017)

In the early part of this century, a group of researchers made the first attempt at producing a scientifically based triage system (Cone & MacMillan, 2005), the Sacco Triage Method (Sacco, Navin, Fiedler, Waddell II, Long, & Buckman, 2005). The resulting system relied on “RPM”, respiratory rate, pulse rate, and best motor response, and scored patients in a manner similar to the Glasgow Coma Score. This system produces a score from zero to 12, and patients are divided into three score bands based on predicted survivability over time. The system has not seen wide use, despite being simple to perform, and with backing literature to show that their selection of RPM was nearly as good a predictor of patient survivability as the Trauma Score, while being faster to assess.

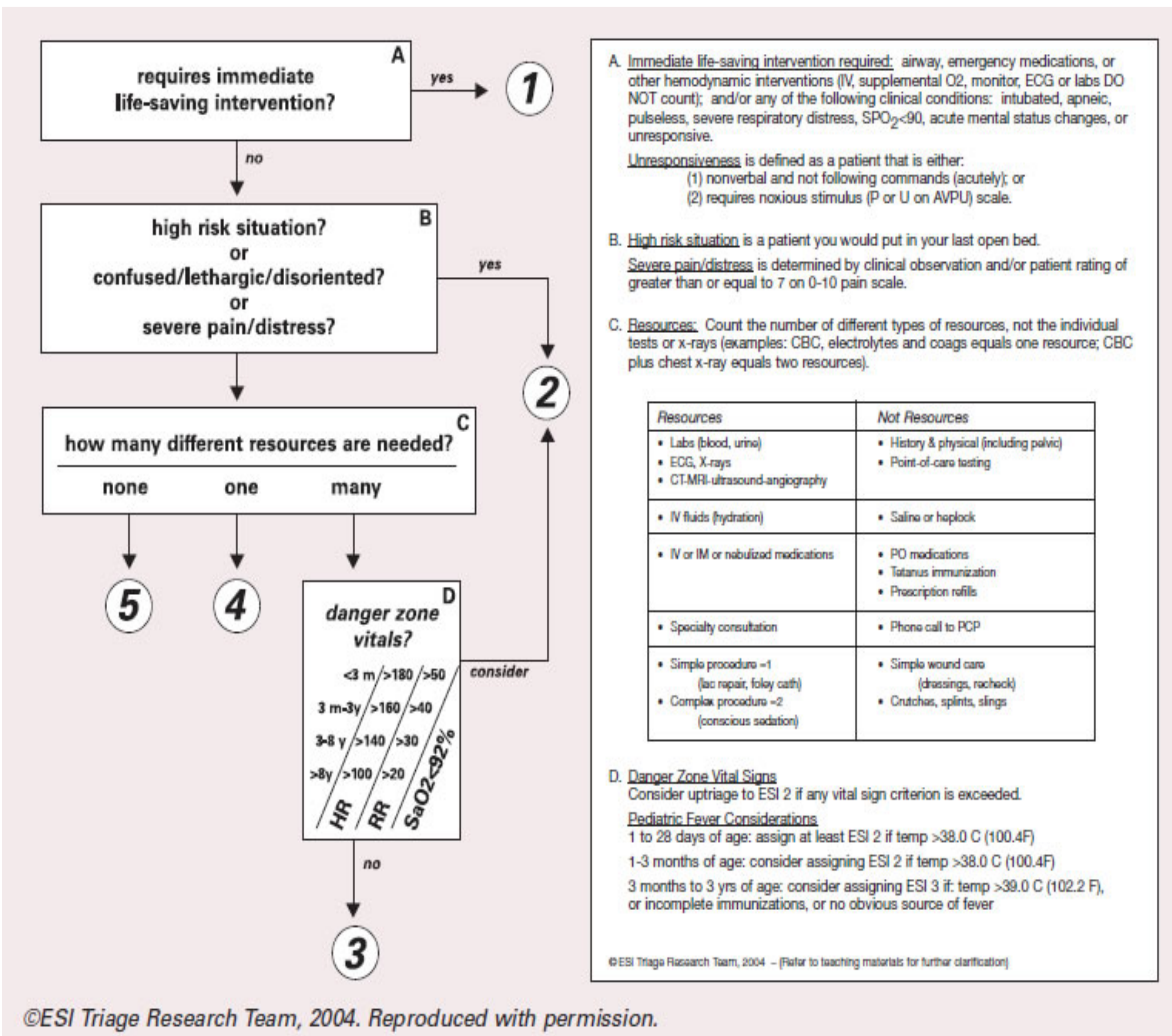

Figure 2: ESI Triage Algorithm. (Agency for Healthcare Research and Quality, 2014)

As noted above, Lerner, et al. (2008) developed the SALT triage system, in an attempt to settle the debate on the best triage methodology. Their work was likewise science based, but SALT makes no accommodation for the patient's progression over time, and the system is more complex than STM. It does, however eliminate the use of the patient's ability to walk as the first decision point, as was seen in the START algorithm. Lerner, et al. have gone on to further propose the implementation of a national standard for triage methodology, which is discussed further, below. Their system, at this point, shows the most promise of becoming a true national standard for triage, although it has not been evaluated against the more empirical systems, such as STM. The danger with implementing a national standard prior to fully evaluating all systems is that once a system becomes codified in regulation, it will become dramatically more difficult to change to another system, even if that new system can be proven to be definitively more effective. The SALT triage method is shown in Figure 3.

Within the last decade, the proliferation has increased, as a 2014 study was able to identify at least five distinct systems in common usage in the United States (Culley, Svendsen, Craig, & Tavakoli, 2014). Their work showed little difference in any of the systems relative to the outcome of over 630 casualties of a chlorine chemical spill in Graniteville, SC in 2005.

In addition to the systems mentioned above, several unique systems are in use internationally. Canada employs the Canadian Triage and Acuity Scale, Denmark utilizes the Copenhagen Triage Algorithm, and some localities in Australia use the Australasian Triage System. Spain has likewise implemented a unique system, META, which is based solely off of Advanced Trauma Life Support (ATLS) protocol (González, et al., 2016). This system has some merit in that it employs a two-stage triage method, wherein immediate life threats are stabilized initially and then those patients are pushed directly to evacuation while less severely injured patients receive additional triage and treatment in the field prior to transport. It is, however, limited by the fact that it adheres strictly to the ATLS protocol, which has been shown to have significant limitations in the management of penetrating trauma (Butler, Hagmann, & Butler, 1996), which led to the development of the newer Tactical Combat Casualty Care (TCCC) protocol. An additional criticism of this system is that it does not make use of an expectant category. A system that does not afford the ability to identify patients with little chance of survival is guaranteed to lead to a misdirection of resources, at some level.

Of note regarding the matter of the use of ATLS or TCCC protocols, this writer is only aware of one study evaluating a triage system against both civilian and battlefield injury. Eastridge, et al., (2010) tested their Field Triage Score (FTS) system against a retrospective selection of patients injured in Afghanistan between 2002 and 2008. They state that their results validate their system against battlefield trauma patterns as well as the civilian trauma patients used a previous study. The older study does not appear to be available for review, and is incompletely cited in Eastridge, et al. (2010), further inhibiting this writer's ability to verify their claims. They did, at least, make an attempt at correlating their results with civilian injury patterns to the ones more commonly seen in a combat arena. This writer could also find no information regarding any agencies that employ the FTS triage system. Its methodology is thoroughly explained in Eastridge, et al. (2010), and so an independent party could conduct a further examination.

While Australia's geographic isolation limits the potential for interagency difficulties arising from the use of multiple triage systems on incidents that occur near international borders, Canada, Spain, and Denmark do not share the same luxury. In the new era of international humanitarian relief efforts, a significant mass casualty incident will draw responders from a global pool, as evidenced by the 2010 earthquake in Port-au-Prince, Haiti. The use of multiple triage systems in the management of a single incident is an invitation for poor patient outcomes and increased mortality.

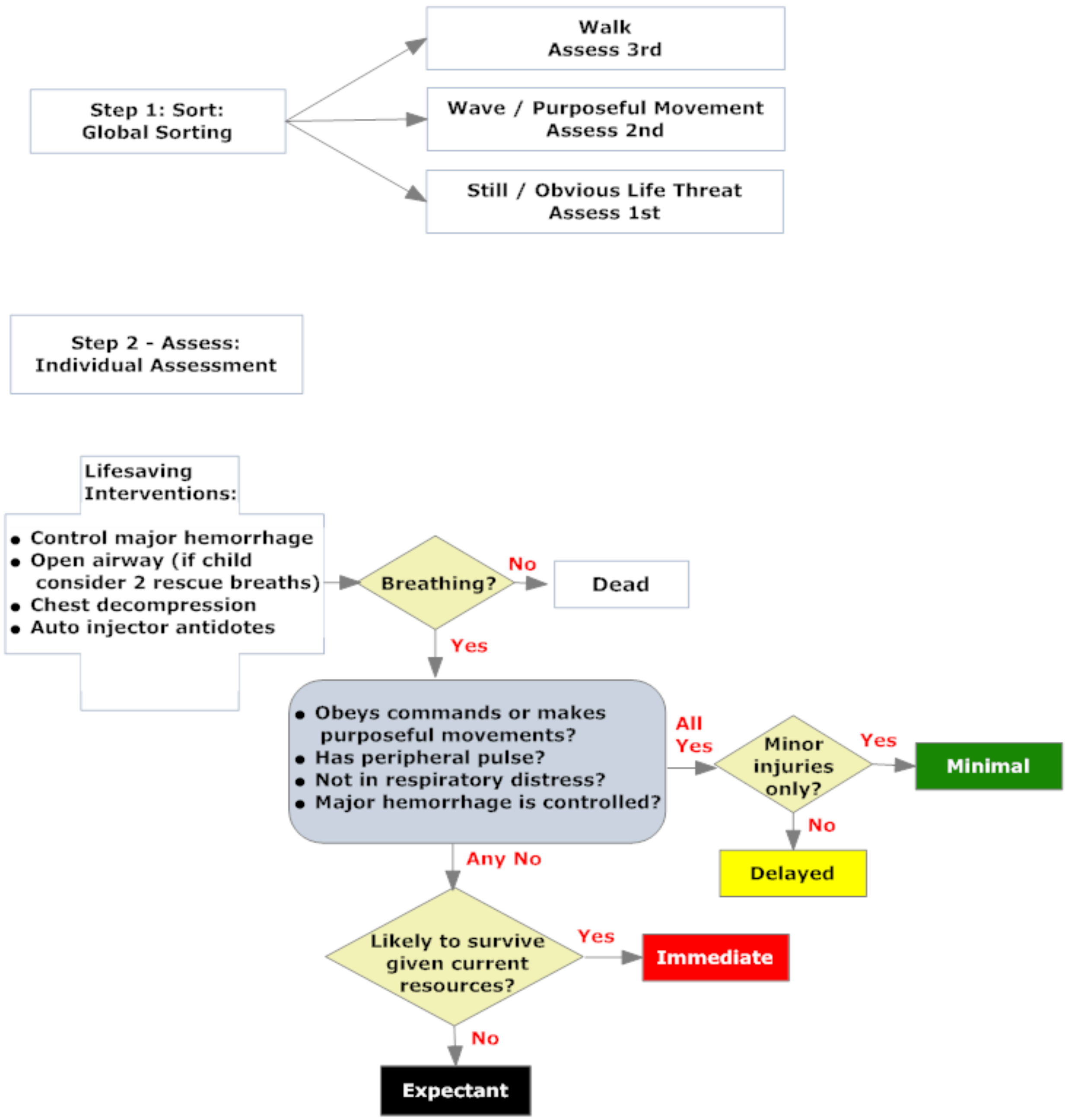

Figure 3: SALT Triage Algorithm. (SALT Triage Algorithm, 2017)

Methodological Problems in the Literature

Among the literature that addresses the effectiveness of the various triage systems, there are a few common methodological problems. The first stems from the previously mentioned lack of a clearly defined standard against which to compare the studied systems. In the absence of such a benchmark for system performance, studies have selected a number of comparison points, from evaluating the individual physiologic markers used by the various systems against patient mortality (Garner, et. al., 2001) to using patient outcomes to determine if the correct triage category was applied by the tested system (Bhalla, Frey, Rider, Nord, & Hegerhorst, 2015). Lerner, et. al. (2015) have provided one benchmark, although its validity has yet to be confirmed by independent analysis. The ISS has also shown to correlate well with patient morbidity and mortality in the mass casualty setting, and at least the START algorithm tracks against the ISS well (Hong, Sierzenski, O'Connor, Bollinger, & Durie, 2007). Given that the ISS has been in use for more than 30 years, it is, at current standing, a more validated surrogate for triage categories than Lerner, et. al.'s classifications.

The second common methodological problem is the consistent use of a field setting triage scenario in the evaluation of the various systems (Khan, et al., 2016) and (Silvestri, et. al., 2017). The underlying premise is that, because of the obvious ethical concerns of testing a novel triage system in an actual mass casualty, researchers opt to simulate a mass casualty incident, and have medical personnel utilize the tested system to categorize the simulated casualties. This method presents a pair of confounding factors. First, in a mass casualty incident, real or simulated, triage personnel must evaluate each casualty in the totality of the presented scenario. This necessarily introduces variables into the triage decision that are not a part of the tested algorithm. An example of this would be a patient who, with available transportation to definitive care, would be classified as "immediate" but, if evaluated *without* available transport for a period of 45 minutes or more, may be categorized into the expectant category, because their wounds are such that they can only be saved with aid not available in the field. The second confounding factor is that, in evaluating a triage system through a field scenario, care must be taken to not provide clues to the patient condition that are not expressly asked for by the triage algorithm in question. If the person providing triage is given clinical data that is not part of the system's decision algorithm, such as a blood pressure, that information will be calculated into the decision making of that triage provider. At that point, whether the patient is properly categorized is no longer a function of the effectiveness of the tested system, but rather of the individual triage provider's clinical acumen.

The final methodological problem to be addressed is specific to Silvestri, et. al.'s 2017 work. That is one of potential bias, the failure to control for such, and the failure to disclose the possibility in the report of their study. Silvestri, et. al. designed their study to utilize the consensus standard put forward by Lerner, et. al. in 2015 as the benchmark against which to evaluate the START and SALT algorithms. In doing so, they may have inadvertently biased the results of their research. The SALT algorithm was originally proposed by Lerner, et. al. in 2008, and then revised (Lerner, et al., 2011). This is potentially problematic, in that eight of the fourteen researchers involved in the development of the 2015 consensus standard were also involved in either the 2008 or 2011 studies which produced and subsequently revised the SALT algorithm tested by Silvestri, et. al. While this should not be taken as impugning the motives of the researchers in question, it does pose a concern to this researcher that a common thought paradigm informed both the creation and revision of SALT and the establishment of the consensus standard. If such is the case, it could artificially skew the results toward the SALT algorithm. To truly validate Silvestri, et. al.'s results, a similar correlation should be sought against a different, but statistically valid benchmark.

This is of particular significance in light of the release of the Model Uniform Core Criteria (MUCC) (2011). This set of 24 criteria has been endorsed by the Federal Interagency Committee on Emergency Medical Services (FICEMS) (2014), and recommended for adoption as a national standard for triage methodologies. The MUCC is intended to not be a triage system in itself, but a set of criteria by which any given triage system can be evaluated. The MUCC was developed by a working group at the Centers for Disease Control and Prevention, and was based on the work of Lerner, et al. (2008). Of note and concern, the FICEMS report states, "Of MUCC's 24 criteria, 15 are currently used by existing MCI triage systems, excluding SALT, which is completely MUCC compliant." At *prima facie*, it appears disingenuous to fail to acknowledge that a significant overlap exists between the research team that initially proposed, and whose work was foundational to, the development of MUCC, the research team that developed the only system in existence that is completely compliant with the proposed Federal guidelines for triage systems, and the research team that developed the 'gold standard' for how a given patient should be categorized by a triage system.

Methodology

Research Theory

This quantitative correlational study will establish a relationship between the three most common triage systems used in the United States, START, SALT, and ESI, and the reference standard provided by the ISS. This approach is selected so that a single sample group, developed through retrospective extraction from the National Trauma Data Bank, can have each of the triage protocols applied to it, and the results of each application correlated with the reference standard. ISS is selected as a reference standard due its ease of use and because it has a strong track record of correlation with patient mortality over the past three decades.

Hypotheses

This study will evaluate four distinct sets of hypotheses. The first two sets will evaluate the correlation of the three tested systems to the reference standard of the ISS. Two hypothesis sets are needed because triage categories can be considered in one of two ways regarding data typology. If one approaches the categories simply as categories, they can be considered nominal data points. In order to test the correlation from this perspective, one must use some form of a chi-square test, in this case for goodness of fit. If, however, one approaches triage categories from the perspective that they represent an increase in injury severity, the data can be viewed as ordinal. For the nominal perspective, the hypothesis is as follows:

$H1_0$: Triage categories produced by the tested algorithm are consistent with the distribution of ISS-based triage categories in the sample.

$H1_a$: Triage categories produced by the tested algorithm are not consistent with the distribution of ISS-based triage categories in the sample.

The dependent variable for this hypothesis is the triage category produced by the tested algorithm, and the independent variable is the triage category indicated by the patient's Injury Severity Score.

For the second hypotheses set, which approaches the data from an ordinal perspective, the hypothesis is:

$H2_0$: There is no statistically significant correlation between the triage category assigned to the patient by the tested system and the triage category indicated by the ISS.

$H2_a$: There is a statistically significant correlation between the triage category assigned to the patient by the tested system and the triage category indicated by the ISS.

The dependent variable is the triage category assigned by the tested algorithm and the independent variable is the triage category indicated by the ISS. This hypothesis will be tested against both the raw ISS and the ISS band based triage category.

The third hypothesis will provide a direct comparison between the accuracy of the three tested systems. A comparison will be made between the percentage of correct assignments from each system, relative to the ISS-based category. For this test, the hypothesis is:

$H3_0$: There is no statistically significant difference in the proportion of triage assignments that agree with the ISS-based category.

$H3_a$: There is a statistically significant level in the proportion of triage assignments that agree with the ISS-based category.

The dependent variable is the proportion of agreement with the ISS-based triage category and the independent variable is the triage system employed. This hypothesis set will be tested for each unique pairing of the tested systems.

The fourth and final set of hypotheses will evaluate whether each triage system is more likely to overtriage or undertriage any given patient. For the purposes of evaluating this hypothesis, overtriage will be defined as the assignment of a triage category that indicates that the patient is more severely injured than the ISS, and undertriage is defined as the assignment of a triage category that indicates the patient is less severely injured than the ISS. This avoids the convolution normally associated with these terms, wherein a patient erroneously assigned to the expectant category, rather than immediate, would be considered a case of undertriage, despite the fact that an expectant patient is more severely injured than an immediate one. Expressly defining these terms in this manner also avoids the pitfall found in Hong, et al. (2008), notably that their report considers a patient undertriaged by being assigned to the Expectant category, but then later considers an Expectant patient to be more acute than an Immediate one, which is key to the linear correlation they found with the ISS. For this portion of the study, the hypothesis is:

$H4_0$: There is no difference in the rate of under- and over- triage, relative to the category indicated by the ISS.

$H4_a$: There is a statistically significant difference in the rate of under- and over- triage, relative to the ISS category.

The dependent variable for this hypothesis is the rates of under- and over-triage, and the independent variable is the triage system employed.

Data Collection and Analysis

To conduct this study, a sample of 180 patients were extracted from the 2016 Research Data Set (RDS) provided by the National Trauma Data Bank (NTDB). The NTDB RDS consists of a series of relational tables, and contains clinical information regarding patients seen at emergency rooms throughout the United States that were assigned a trauma related diagnosis code. Each patient visit is assigned an incident key that is used to relate the information collected across the tables in the database. This incident key also renders the data anonymous in order to ensure the patient's privacy rights are protected. There is no way for a researcher to connect the information contained in the RDS to any individual person, living or dead.

The sample size was selected to achieve a level of significance (α) of 0.05 and a power ($1-\beta$) of 0.95 with an effect size ($|\rho|$) of 0.3. The minimum sample size to meet these testing parameters is 134; 180 was selected as the final sample size in order to facilitate the selection of patient records from the RDS; 45 records were selected for each of four triage categories, using the ISS ranges established in studies of Israeli mass casualty incidents (Kosashvili, Daniel, Peleg, Horowitz, Laor, & Blumenfeld, 2009), ISS of 1-8 were categorized as "minimal", or "green", 9-24 as "delayed" or "yellow", 25-74 as "immediate" or "red", and a score of 75 indicated the "expectant" or "black" category. The selection of 45 patients from each ISS band also ensured the evaluation of the ESI system would meet the minimum sample size, for reasons explained below.

A four-tiered system was chosen to create the most analogous category scheme across the three systems. START uses the four tiers listed. SALT uses a five-tiered scheme, but differentiates dead casualties from those merely expected to expire, unlike START. For the purposes of this study, results of "dead" or "expectant" from the SALT algorithm were combined into one category. ESI, likewise, uses a five-tiered scheme, but can really be viewed as a three-tiered scheme, with category 1 corresponding to "immediate", category 2 to "delayed", and category 3, 4 or 5 all corresponding to "minimal", as the system considers these patients to lack the acuity needed to be treated urgently in the Emergency Department (ED) setting. Notably this system does not have a category for expectant patients; the assumption behind this is that once the patient arrives at the ED, they will receive treatment rather than being relegated to merely palliative care. In light of this, patients with an ISS of 75 were not evaluated using ESI, since all 45 of them would be incorrectly categorized, artificially skewing the results against this system.

For each of the tested triage algorithms, the answers to the decision points will be extracted for each patient. This required several assumptions on the part of the researcher, as many of the questions asked in the algorithms are not directly answered by the RDS. The following describes the method by which the patients were extracted from the RDS and the algorithm answers were generated.

Extraction of patients from the RDS. The RDS contains data on more than 1.6 million patients seen in the ER and assigned a trauma diagnosis code. The process began with the table titled RDS_ED, which contains information regarding the Emergency Department visit, Injury Severity Score, and discharge status of the patients in the RDS. This table was filtered down to contain only the Incident Key, the discharge type from the ED, whether the patient arrived at the ED with signs of life, and the Injury Severity Score. This table was then related to another table, titled RDS_VITALS, based on the Incident Key. This provided the vital signs information for each of the patients in the RDS_ED table, as recorded by Emergency Medical Services in the field. This data was then filtered to eliminate entries that contained unusable data values in any field on the table. The resulting table was then related to a third table from the RDS, titled RDS_ICD10_DCODE. This relation provided every diagnosis code for each patient entry in the data set. Because of the formatting of this last table, the total number of entry rows on the data set exceeded 9 million at this point. The data was then sorted according to Incident Key and the first 195,000 entries were extracted to bring the file to a more manageable size. This selection was then sorted according to Injury Severity Score, and 45 Incident Keys were selected at random from each of the ISS bands listed above, distributed across the ISS band. Lastly, these 180 incident keys were related to the table RDS_DEMO, which provided the age of each patient. The final selections were then evaluated to produce a master data table containing the answers to the triage algorithm questions according to the process and assumptions described below.

Generation of algorithm answers. Beginning with the START algorithm, the first data point to be extracted was whether the patient was able to walk. This necessitated the first assumption. A patient was determined to be non-ambulatory if their record contained a diagnosis code that indicated an open wound, fracture, or amputation at any point below the waist, any cervical spine fracture, any major thoracic trauma, or a Glasgow Coma Scale (GCS) of 12 or less. The next step of the START triage process is to evaluate whether the patient is breathing spontaneously, and if not, whether opening the airway was successful at restoring spontaneous respirations. The RDS contains no information by which the researcher could determine if a patient had spontaneous respirations restored by opening the airway, and so these two questions were evaluated as one. It is possible, as a result of this assumption, that a patient could have continued all the way down the algorithm and been triaged as “delayed” when they should have been categorized as “immediate” but this is a limitation of the data set, and could not be avoided. Based on the resulting data, it is unlikely that this possibility had any significant impact on the final result of the study, as will be explained later in the report.

The next step in START is to evaluate whether the respiratory rate is greater than or less than 30. Patients with a respiratory rate of exactly 30 were included with those with a rate greater than 30. START then evaluates for the presence or absence of a radial pulse, or whether the capillary refill time is greater than or less than two seconds. The RDS does not provide information regarding the patients’ capillary refill time; patients were deemed to have a palpable radial pulse based on the widely accepted correlation to a systolic blood pressure of 90 or greater. The final question asked in the START algorithm addresses the mental status of the patient. This was evaluated off of the GCS Motor component score; any patient with a 5 or lower on the GCS Motor score was deemed to be unable to obey commands.

Moving to the SALT algorithm, the study did not incorporate the Global Sorting component of the system. This is because this step does not have the potential to alter the final triage category assignment, merely the order in which each patient will be evaluated using the individual assessment. The first question in the assessment is whether the patient is breathing;

this was answered based on the respiratory rate recorded by EMS in the field. The next step in the algorithm addresses four separate questions, all of which must be answered yes to result in categorization as either “minimal” or “delayed”. The first of these questions is whether the patient makes purposeful movements. This was answered by again turning to the GCS Motor score; any patient with a motor score of three or higher (meaning that they had retained the ability to withdraw appropriately from a painful stimulus) was considered a yes answer. The next question addresses the presence of a peripheral pulse, again with a systolic blood pressure of 90 or greater indicating a yes answer. Then the algorithm asks whether the patient is not in respiratory distress. A patient with a respiratory rate outside of normal physiological limits (12-20 respirations per minute) or with a pulse oximetry of 93% or less was deemed to be in respiratory distress. Care was taken with this question to note the intentional double negative in the wording and provide the correct answer relative to the algorithmic tree. The final question in this step is whether or not major hemorrhage has been controlled. This was answered by again referring to the systolic blood pressure; any patient with a systolic blood pressure of 90 or higher was deemed to have any major hemorrhage successfully controlled.

This step in SALT divides patients by whether all four questions were answered “yes” or at least one question was answered “no”. Patients with all “yes” answers are then evaluated on whether they have minor injuries only, or not. Patients with an ISS of nine or greater were considered to have more than minor injuries. Patients with at least one “no” answer in the second step are evaluated on whether or not they are likely to survive with the current resources available. This highly subjective question was answered based on whether the patient arrived at the ED exhibiting signs of life, indicating their ability to survive with the resources contained within a single ambulance.

Now evaluating the ESI method, the first question asked is if the patient requires a lifesaving intervention. This was answered in the affirmative for patients with a diagnosis code of an obvious life threatening injury, any vital sign significantly outside the normal physiological range, or a total GCS of eight or below, the commonly accepted threshold for endotracheal intubation. The next step in the ESI algorithm asks three questions: is the patient in a high-risk situation, are they confused, lethargic, or disoriented, and are they in severe pain or distress. Patients were considered to be high risk if their ISS was 25 or greater and any patient with a Glasgow Coma Scale below 15 was considered to be confused, lethargic, or disoriented. Severe pain or distress was evaluated by a pulse of 100 or higher in the absence of a systolic blood pressure below 90, and a respiratory rate greater than 22 without a pulse oximetry of 93 or below.

The RDS does not provide information that could easily determine the number of resources needed to treat the patient in the ED, but this question was deemed unnecessary for evaluation as it merely provides a differentiation between categories 3, 4, and 5, all of which were considered minimal for the purposes of this study. The final question is whether the patient's vital signs fall into the delineated "danger zone" based on age. ESI directs the person conducting triage to consider elevating patients with vitals in this danger zone to category 2. This is consistent with the common adage in medicine to "treat the patient, not the vital signs". In the absence of an actual patient to evaluate, any patient with a yes to this question was elevated to category 2.

Data collection. Once the master data file was completed, the patient list was sorted according to Incident Key, and numbered from 1 to 180. Each algorithm was then separated into an individual datasheet. These three datasheets, along with a set of instructions and a copy of one of the three studied algorithms, were provided to volunteers from the military staff of the Branch Health Clinic, Naval Station Norfolk. These volunteer handouts may be found in Appendices A, B, and C. The use of volunteers to run the patient data through the three studied algorithms was included to minimize the potential for bias on the part of the researcher. No data was collected regarding the identity of these volunteers, as they and their triage performance was not the subject of this study. The datasheets contained the answers to all of the questions asked by the algorithm, regardless of where the individual patient might stop in the algorithm. Volunteers were instructed to use the information in the datasheet to process each patient through the algorithm they had been provided, and record the first triage category that their algorithm produced. Each patient was processed one time through each of the three studied algorithms, with the exception of the 45 patient records with an ISS of 75, which were only processed through the START and SALT algorithms. These patient entries were greyed out on the handouts for the ESI algorithm to ensure that results were not misapplied to a different patient number.

The H1 hypothesis set was evaluated using a free online statistics calculator (Stangroom, 2018). To evaluate the H1 hypotheses, the total count of each triage category was compared to the expected distribution based on the total number of patients selected for each category. A Chi Square test for goodness of fit was then conducted to determine how closely the triage results matched the expected distribution of category assignments. This test allows a comparison of the sample distribution to the expected distribution in order to evaluate whether any deviation from the expected distribution is due to a statistically significant factor (in this case the efficacy of the individual triage system employed), or mere random variation. Since the patient records used in the study were selected specifically in accordance with a fixed distribution, this test provides a strong indicator of whether the tested system produces correct results.

Mathematical analysis of the H2 hypothesis was conducted using another online calculator (Vasavada, 2016). The H2 hypotheses were evaluated by assigning both the triage category results and the reference triage category a numerical ranking, increasing in order of injury severity. Minimal was coded as “1”, Delayed as “2”, Immediate as “3” and Expectant (inclusive of patients categorized as “dead” in the SALT algorithm) as “4”. This assignment of a numerical ranking allowed for a quantitative examination of any correlation between an increase in the ranking of the triage assignment produced by the tested system and increasing Injury Severity Scores. Plotting the resulting category pairs on a scatter diagram initially did not display the data in a useful way, but a marked line chart, when data was sorted according to the ISS-based category, was found to reveal trends adequately. Correlation was established via the use of a Kendall *tau* test to determine the sample correlation coefficient for each algorithm. This test method was selected due to the fact that in evaluating the data from the ordinal perspective, accommodation must be made for tied data. This test series was conducted twice; once evaluating the correlation between the assigned triage category and the ISS-based category, and once to examine the correlation between the assigned triage category and the raw ISS score.

The H3 hypothesis set was again evaluated using Stangroom’s online calculator (2018). Evaluation of the H3 hypothesis was tested by first determining the number of cases of agreement with the ISS-based category for each tested system to establish the sample proportion. A two-tailed difference of two proportions z-test was then conducted for each unique pairing of systems; START versus SALT, START versus ESI, and SALT versus ESI. Comparison of the results was then evaluated to rank the three systems in relation to each other. This test was conducted as a two-tailed rather than right or left tailed test in order to evaluate for any statistically significant difference between the two proportions, positive or negative. The rates of agreement were also plotted on a clustered bar graph to provide a visual comparison between the two systems.

The H4 hypothesis set was also conducted using Stangroom's calculator (2018). To evaluate the H4 hypotheses, for each triage system, the assigned category was compared to the ISS-based category and the number of cases of agreement or disagreement was noted. The cases of disagreement were then determined to either be a case of undertriage or overtriage, as defined above, relative to the ISS-based category. A binomial probability was calculated for each system, noting the probability of producing a number of over- or under-triage cases less than or equal to the noted results, with a number of trials equal to the number of disagreements between the assigned category and the ISS-based category. Examining the data with this test allows us to evaluate the probability of the tested system producing up to the number of cases of over- or under-triage noted in the sample. This test serves to evaluate the likelihood that the tested system would produce the proportions of over- and under- triage used in the next step of evaluating this hypothesis. A two-tailed difference of proportions z-test was then conducted to estimate the population proportions of overtriage and undertriage relative to the total disagreement sample. These data were then plotted on a clustered bar graph. One bar represented overtriage, a second represented undertriage, and a final bar represented the number of cases of agreement. Figure 4 presents all hypotheses, data sources, and analyses conducted in a tabular format.

Hypothesis	Data Sources Used	Statistical Analysis Conducted
H1: Evaluating conformity to expected patient distribution	Sample distribution and expected distribution.	Chi Square test for Goodness of Fit
H2: Evaluating correlation between system produced triage assignment and the Injury Severity Score	Numerically ranked triage categories produced by the tested systems and the ISS-based triage category	Kendall *tau* correlation test
H3: Evaluating the probability of the tested systems agreeing with the ISS-based category	Count of cases of agreement with the ISS-based category for each tested system and the total sample size	Two-tailed difference of two proportions *z*-test
H4: Evaluating the probability of the tested systems over-or under-triaging a patient relative to the ISS-based triage category	Counts of over- and under- triage and counts of disagreement with the ISS-based triage category for each tested system	Binomial probability test and two-tailed difference of two proportions *z*-test

Figure 4: Hypotheses, Data Sources, and Analyses Conducted

Results

Hypothesis One

This hypothesis evaluated whether the triage assignments produced by the tested systems would conform to the expected distribution. In this study, there were 45 patients selected for each of the ISS-based triage category bands. $H1_0$ was that the triage assignments produced by the tested systems would conform to the expected distribution, and $H1_a$ was that the triage assignments produced by the tested systems would not conform to the expected distribution. For testing this hypothesis, the dependent variable was the count of triage assignments to each category for each tested system and the independent variable was the expected distribution of 45 assignments per category.

START triage. The START algorithm was found to have produced 64 Minimal assignments, 39 Delayed assignments, 52 Immediate assignments, and 25 Expectant assignments. The expected distribution was 45 for each of the four categories. Calculation of the Chi Square at $\alpha = 0.05$ yielded a result of $x^2 = 18.8$ with a P-value < 0.001. The P-value is ≤ 0.05, and so we must reject the null hypothesis. The START algorithm did not fit the expected distribution.

Figure 5, below, displays the count data for the START algorithm in a clustered bar graph. Each clustered pair of bars represent the expected and observed counts for one of the four triage categories. The blue bars represent the expected distribution and the red bars represent the actual count of assignments to that triage category by the START algorithm. The vertical axis represents the count, and there is no specific quantitative data point on the horizontal axis. Moving left to right along this axis, the triage categories are arranged in decreasing patient acuity.

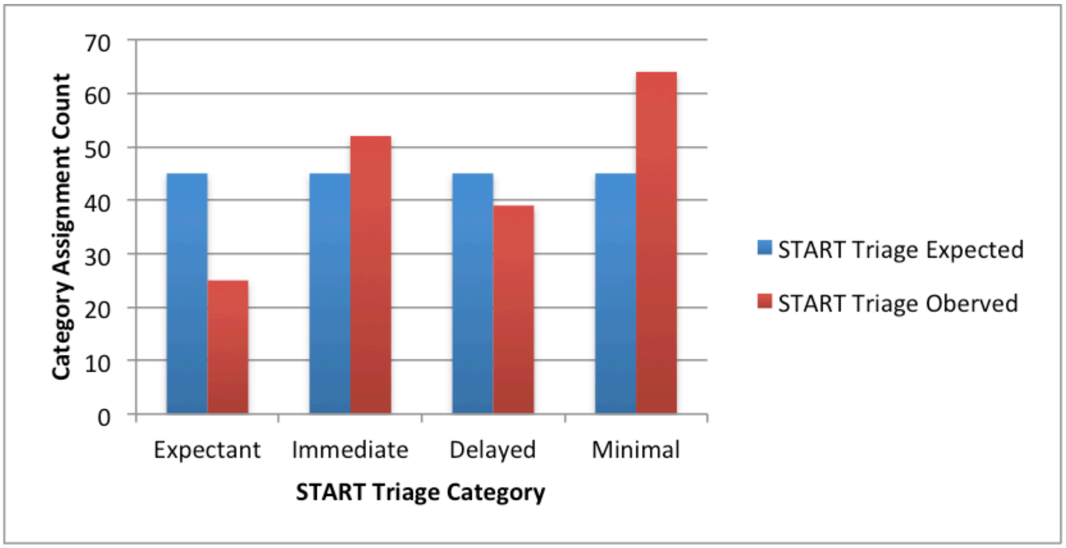

Figure 5: START Triage Goodness of Fit

SALT triage. The SALT algorithm resulted in 46 Minimal assignments, 54 Delayed assignments, 53 Immediate assignments, and 27 Expectant assignments. The expected distribution was 45 for each of the four categories. Calculation of the Chi Square at $\alpha = 0.05$ yielded a result of $x^2 = 10.44$ with a P-value < 0.015. The P-value is ≤ 0.05, and so we must reject the null hypothesis. The SALT algorithm did not fit the expected distribution.

Figure 6, below, displays the count data for the SALT algorithm in another clustered bar graph. Each clustered pair of bars represent the expected and observed counts for one of the four triage categories. The blue bars present the expected distribution and the red bars display the actual count of assignments to that triage category by the SALT algorithm. The vertical axis represents the count, and there is no specific quantitative data point on the horizontal axis. Moving left to right along this axis, the triage categories are arranged in decreasing patient acuity. It is worth noting that, of the three tested systems, SALT most nearly conformed to the expected distribution. This point will be explored further in the Discussion section below.

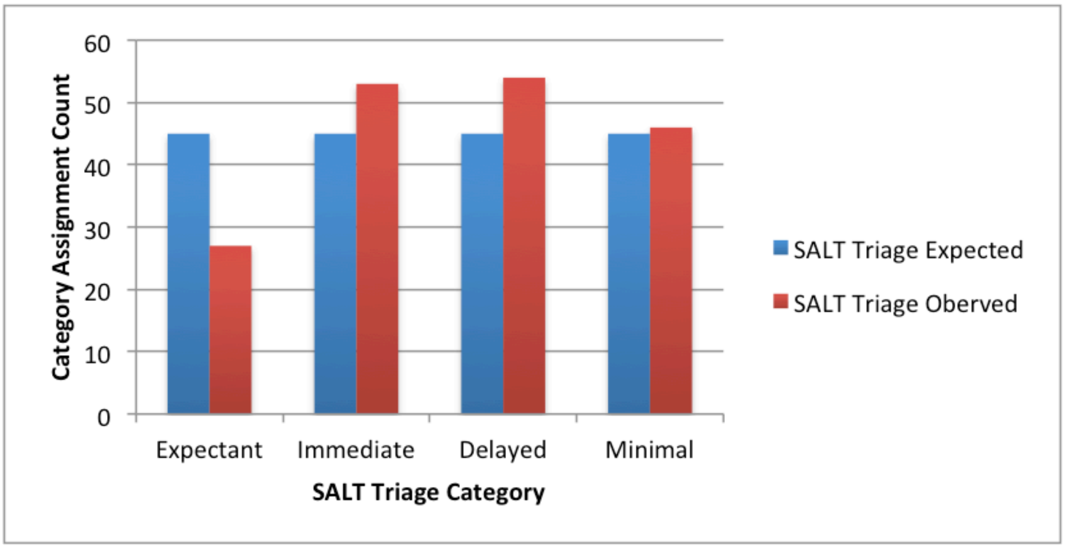

Figure 6: SALT Triage Goodness of Fit

ESI triage. The ESI algorithm produced 33 Minimal assignments, 69 Delayed assignments, and 33 Immediate assignments. The expected distribution was 45 for each of the three categories. Calculation of the Chi Square at $\alpha = 0.05$ yielded a result of $x^2 = 19.20$ with a P-value < 0.001. The P-value is ≤ 0.05, and so we must reject the null hypothesis. The ESI algorithm did not fit the expected distribution.

Figure 7, below, likewise displays the count data for the ESI algorithm in a clustered bar graph. Each clustered pair of bars represent the expected and observed counts for one of the three triage categories applicable to the ESI algorithm. The blue bars show the expected distribution and the red bars represent the actual count of assignments to that triage category by the ESI algorithm. The vertical axis represents the count, and there is no specific quantitative data point on the horizontal axis. Moving left to right along this axis, the triage categories are arranged in decreasing patient acuity. Examining the observed distribution of assignments we see that the amount of patients by which ESI exceeded the expected distribution for Immediate assignments very closely matches to the combined amount of patients by which ESI did not meet the expected assignment count.

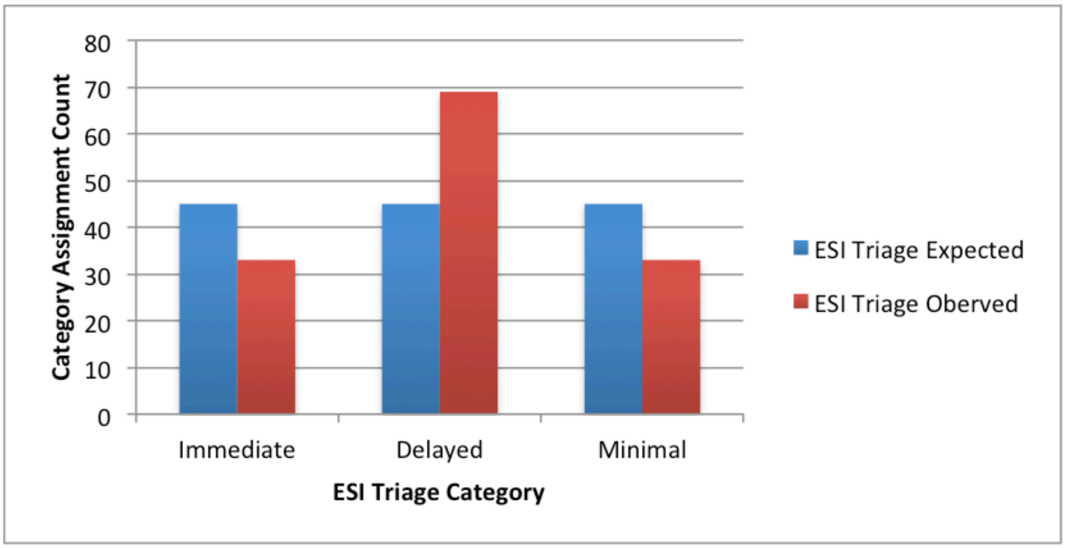

Figure 7: ESI Triage Goodness of Fit

Summation of results for hypothesis one. Regarding the conformity of the observed results for each of the three tested triage systems to the expected distribution, we find that we reject the $H1_0$ in each case. Of the three systems, only SALT produced results that were close to that needed to not reject the null. This will be examined in more detail in the Discussion section.

Hypothesis Two

The second hypothesis evaluated whether the numerically ranked triage assignments produced by the tested systems would correlate to the ISS-based triage category, and also to the raw ISS itself. $H2_0$ was that there is no statistically significant correlation between the triage assignments produced by the tested system and the ISS, and $H2_a$ was that there is a statistically significant correlation between the triage assignments produced by the tested system and the ISS.

For testing this hypothesis, the dependent variable was the triage assignment produced by the tested system and the independent variable was the triage category indicated by the Injury Severity Score.

START triage. A Kendall *tau* test was conducted evaluating the correlation between the assigned category and the ISS-based category. This test resulted in $\tau = 0.5742$, with a P-value of 0.0000. At the $\alpha = 0.05$ confidence level, we reject the null hypothesis that there is no correlation between the triage category produced by the START algorithm and the Injury Severity Score. Repeating the test, but this time comparing the results of the START algorithm to the raw ISS resulted in $\tau = 0.5563$, again with a P-value of 0.0000. At the $\alpha = 0.05$ confidence level, we reject the null hypothesis when the results of START are compared directly to the Injury Severity Score.

Figure 8 displays the correlational data for the START algorithm in a stacked, marked, line graph. Displaying the data from this test in a traditional scatter diagram did not produce a valuable representation of the results as a result of the limited number of values for both the vertical and horizontal axes relative to the total number of data points. Essentially, displaying the data in that way would produce a series of stacked data points the intersections of each pairing of the START triage assignment and the ISS-based category. Using a stacked line, and sorting the results by the ISS-based category reveals the correlation of the START triage assignments to the trend line of the ISS-based category. In this chart, the blue line shows the triage assignments produced by the START algorithm, with each marked data point representing a specific patient. The green line represents the ISS-based triage category for that same patient. The vertical axis represents the numerical ranking of the triage categories in increasing patient acuity from one to four, and horizontal axis shows the specific patient number for the data points at that intersection, from one to 180. The horizontal axis labels are not in numerical order.

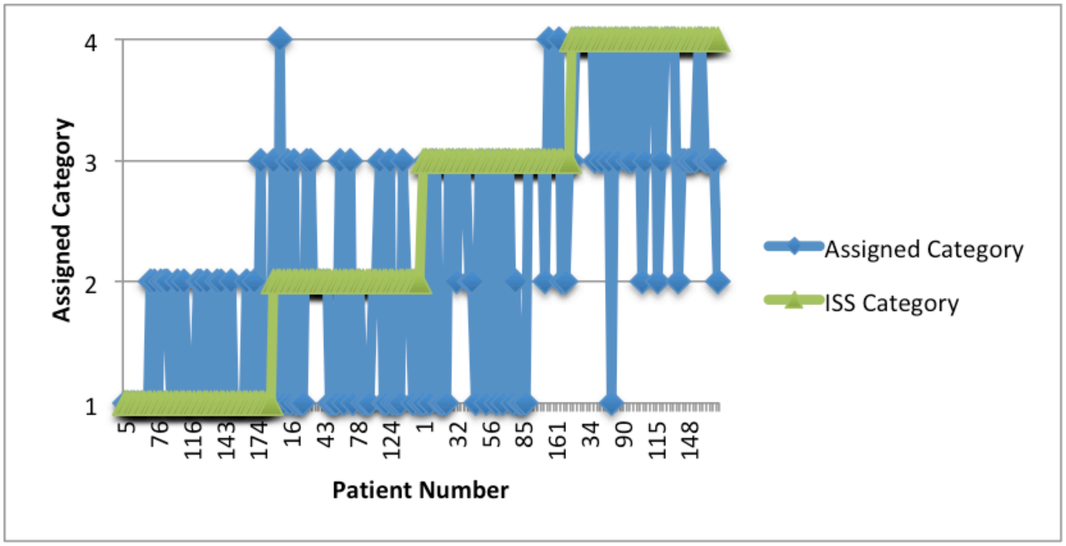

Figure 8: START Triage Correlation

SALT triage. A Kendall *tau* test resulted in $\tau = 0.7430$, with a P-value of 0.0000. At the $\alpha = 0.05$ confidence level, we reject the null hypothesis that there is no correlation between the triage category produced by the SALT algorithm and the Injury Severity Score. Repeating the test, but this time comparing the results of the SALT algorithm to the raw ISS resulted in $\tau = 0.6870$, again with a P-value of 0.0000. At the $\alpha = 0.05$ confidence level, we reject the null hypothesis when the results of SALT are compared directly to the Injury Severity Score.

Figure 9 displays the correlational data for the SALT algorithm in a stacked, marked, line graph. Displaying the data from this test in a traditional scatter diagram did not produce a valuable representation of the results as a result of the limited number of values for both the vertical and horizontal axes relative to the total number of data points. The results of this test produced the same difficulty in the use of a traditional scatter diagram, and the chart shown is organized identically to Figure 8. Using a stacked line, and sorting the results by the ISS-based category reveals the correlation of the SALT triage assignments to the trend line of the ISS-based category. In this chart, the blue line shows the triage assignments produced by the SALT algorithm, with each marked data point representing a specific patient. The green line represents the ISS-based triage category for that same patient. The vertical axis represents the numerical ranking of the triage categories in increasing patient acuity from one to four, and horizontal axis shows the specific patient number for the data points at that intersection, from one to 180. The horizontal axis labels are not in numerical order. Examining the chart shows that the SALT algorithm displays a much closer relationship to the ISS-based category trend line than was seen with the START algorithm.

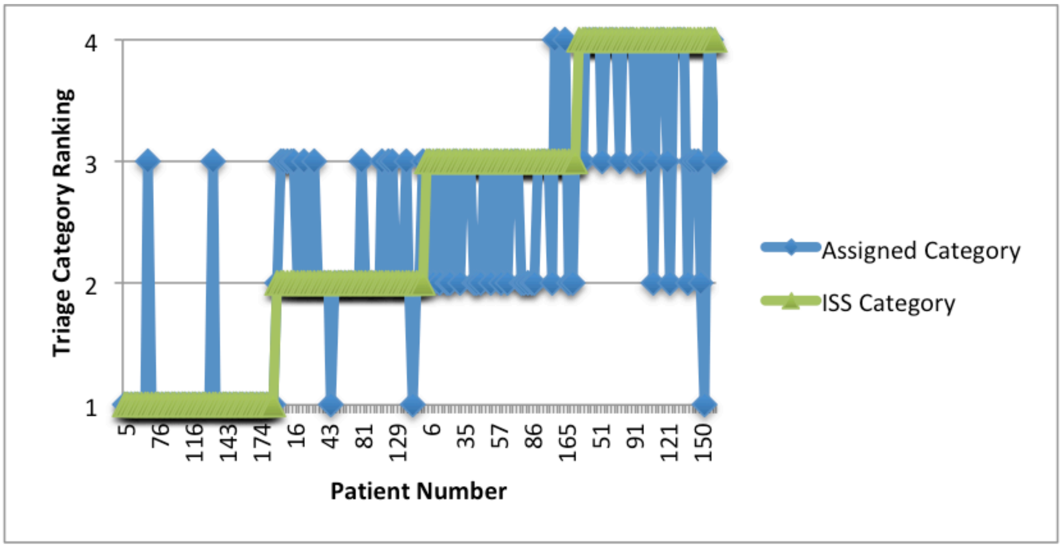

Figure 9: SALT Triage Correlation

ESI triage. The Kendall *tau* test produced $\tau = 0.3874$, with a P-value of 0.0000. At the $\alpha = 0.05$ confidence level, we reject the null hypothesis that there is no correlation between the triage category produced by the ESI algorithm and the Injury Severity Score. Repeating the test, but this time comparing the results of the ESI algorithm to the raw ISS resulted in $\tau = 0.3765$, again with a P-value of 0.0000. At the $\alpha = 0.05$ confidence level, we reject the null hypothesis when the results of ESI are compared directly to the Injury Severity Score.

Figure 10 displays the correlational data for the ESI algorithm in another stacked, marked, line graph. Even though there were a smaller number of patients examined with the ESI system, these results also presented similar difficulties in the use of a scatter diagram. Like Figures 8 and 9 above, the data shown is sorted by ISS-based category to produce a trend line. In this chart, the blue line shows the triage assignments produced by the ESI algorithm, with each marked data point representing a specific patient. The green line represents the ISS-based triage category for that same patient. The vertical axis represents the numerical ranking of the triage categories in increasing patient acuity from one to three, and horizontal axis shows the specific patient number for the data points at that intersection, from one to 135. The horizontal axis labels are not in numerical order. Examining the chart shows that the ESI algorithm results bear much less of a relationship to the ISS-based category than do the results of START or SALT.

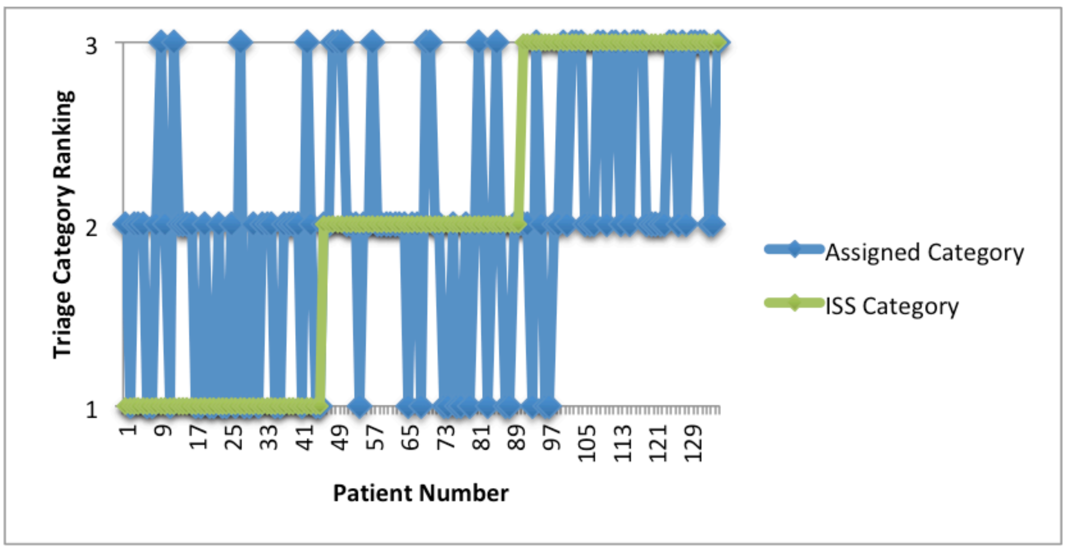

Figure 10: ESI Triage Correlation

Summation of results for hypothesis two. Examining the results of the three tested systems for a correlation with the ISS-based triage categories, we find that in all three cases we reject the null hypothesis. The relative strength of correlations found were SALT, then START, then ESI, in descending order. In all three tests, the P-value was essentially zero, indicating that there is a virtual certainty that we are not making a Type I statistical error.

Moving to correlation between the results of the tested algorithms and the raw ISS, we again reject the null hypothesis for all three systems, although the strength of correlation in each was slightly reduced. The P-value remained essentially zero, however, again indicating that there is very little chance that we are incorrectly rejecting the null hypothesis.

Hypothesis Three

In hypothesis three we evaluated whether the tested triage systems displayed any statistically significant difference in the rate at which they agreed with the ISS-based category assignment. $H3_0$ was that there was no statistically significant difference in the rate of agreement with the ISS-based category, and $H3_a$ was that there was a statistically significant difference in the rate of agreement with the ISS-based category. The dependent variable was the proportion of agreement with the ISS-based category and the independent variable was the specific triage system employed. This hypothesis was tested for each unique pairing of triage systems, START vs. SALT, START vs. ESI, and ESI vs. SALT.

START versus SALT. The START algorithm agreed with the ISS-based category 87 times out of 180 trials for a proportion of 0.483. SALT produced 121 agreements in 180 trials, a proportion of 0.672. A two-tailed z-test produced a Z-score of -3.6281 with a P-value of 0.00028. Comparing START to SALT, at the $\alpha = 0.05$ confidence level, we reject the null hypothesis that there is no difference in the agreement to the ISS-based category.

Figure 11 displays the proportions of agreement with the ISS-based category for the START and SALT algorithms in a clustered bar graph. In this chart, START is represented by the blue bar and SALT is represented by the red bar. The vertical axis represents the proportion of agreement with the ISS-based category, from 0 to 1, labeled in 0.1 increments. There is no specific quantitative element displayed on the horizontal axis. The SALT algorithm agreed with the ISS-based category much more frequently than did START.

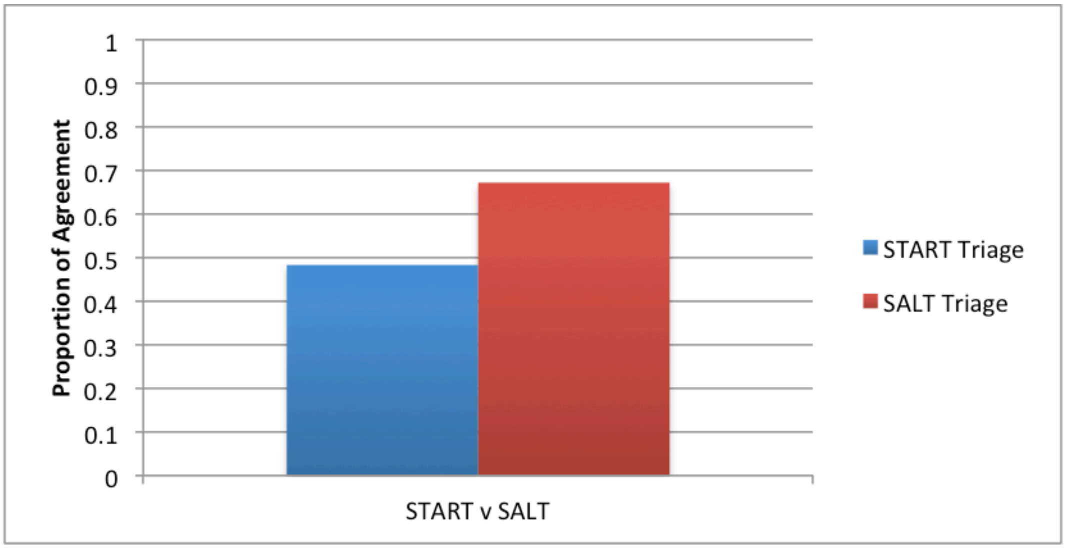

Figure 11: START versus SALT Proportions of Agreement

START versus ESI. ESI agreed with the ISS-based category in 66 of 135 trials. The proportion of agreement is 0.489, compared against the agreement proportion of START, 0.483. A two-tailed z-test yielded a Z-score of -0.0976 and a P-value of 0.92034. When considering START against ESI, at the $\alpha = 0.05$ confidence level, we do not reject the null hypothesis. There is no statistically significant difference in the rate of agreement with the ISS-based category for these two systems.

Figure 12 displays the proportions of agreement with the ISS-based category for the START and ESI algorithms in a clustered bar graph. In this chart, START is represented again by the blue bar and ESI by the green bar. The vertical axis represents the proportion of agreement with the ISS-based category, from 0 to 1, labeled in 0.1 increments. There is no specific quantitative element displayed on the horizontal axis. The rates of agreement with the ISS-based category between the START and ESI algorithms are found to be virtually identical.

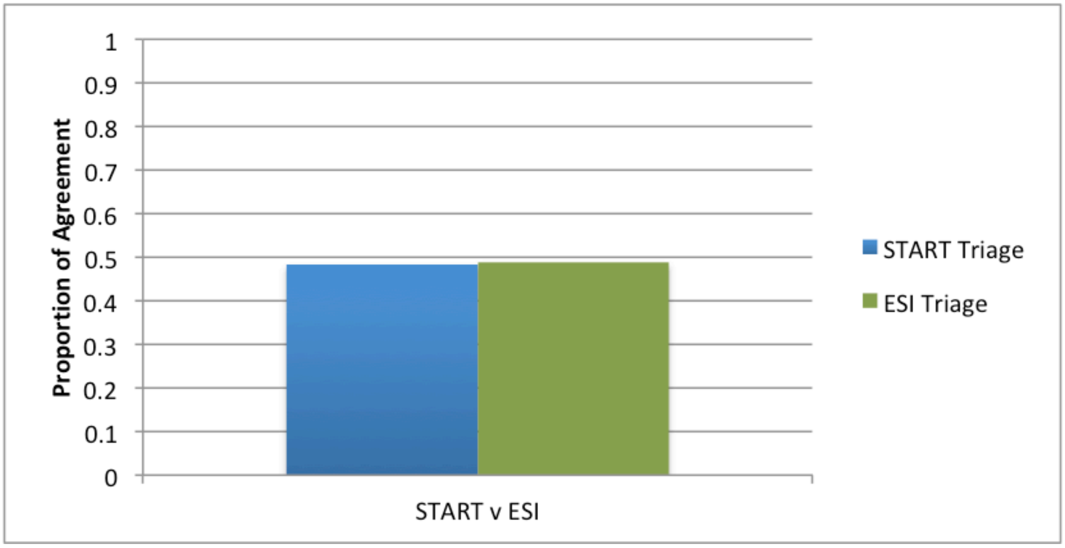

Figure 12: START versus ESI Proportions of Agreement

ESI versus SALT. The agreement proportions of ESI and SALT are as noted above, 0.489 and 0.672, respectively. The results of the two-tailed z-test were a Z-score of 3.2785 and a P-value of 0.00104. Regarding ESI versus SALT, at the $\alpha = 0.05$ confidence level, we reject the null hypothesis.

Figure 13 displays the proportions of agreement with the ISS-based category for the ESI and SALT algorithms in a clustered bar graph. In this chart, ESI is represented by the green bar and SALT is once again represented by a red bar. The vertical axis represents the proportion of agreement with the ISS-based category, from 0 to 1, labeled in 0.1 increments. There is no specific quantitative element displayed on the horizontal axis. We again see that the SALT algorithm results in a much higher rate of agreement with the ISS-based category than does the ESI algorithm.

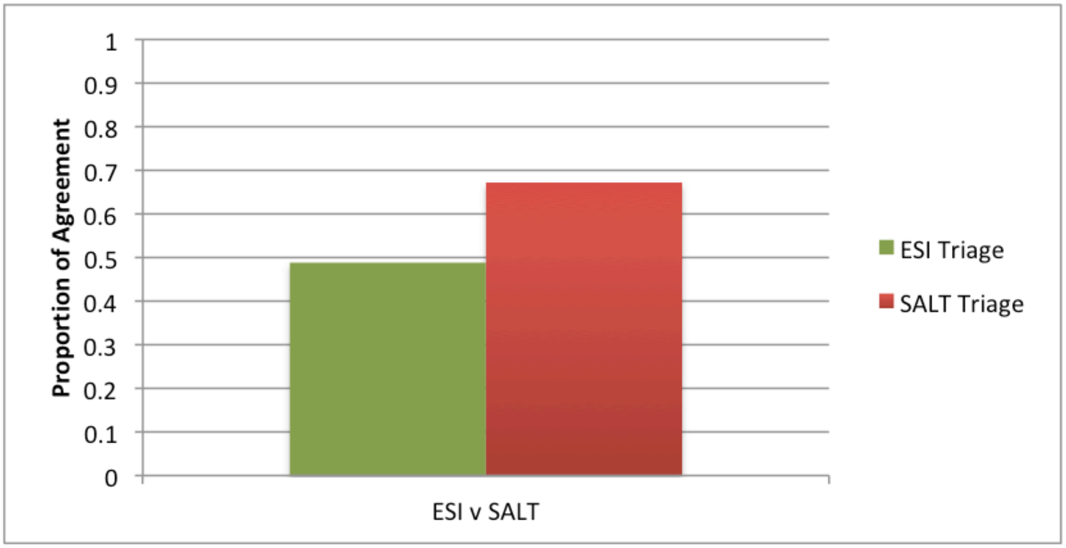

Figure 13: ESI versus SALT Proportions of Agreement

Summation of results for hypothesis three. Comparing START to SALT, we reject the null hypothesis that there is no difference in the rates of agreement between the two systems. Despite testing at $\alpha = 0.05$, the results would have been significant at $\alpha = 0.001$, so there is little likelihood of a Type I error. Comparing START to ESI, we do not reject the null hypothesis that there is no difference in the rates of agreement between these two systems. The P-value for this test was very high, indicating a low chance of a Type II error. When we compare ESI to SALT, we again reject the null hypothesis that there is no difference in the rates of agreement between these systems. In this case, we again find that while testing at $\alpha = 0.05$, the test would have still been significant at a lower confidence level, in this case, $\alpha = 0.01$.

Hypothesis Four

This final hypothesis evaluated whether the tested triage systems were more likely to overtriage or undertriage a patient (as defined above) when they did not agree with the ISS-based category. $H4_0$ was that there was no statistically significant difference in the rate of over- and under- triage relative to the ISS-based category, and $H4_a$ was that there was a statistically significant difference in the rate of over- and under- triage relative the ISS-based category. The dependent variable was the rates of over- and under- triage for each system and the independent variable was the specific triage system employed.

START triage. The results of the START algorithm agreed with the ISS-based triage category 87 times out of 180 trials. Of the 93 disagreements, 29 were overtriage, representing 31.18% of disagreements, and 64 were undertriage, representing 74.19% of cases. A binomial probability distribution resulted in a probability of 0.0001 that the system would produce 29 or fewer cases of overtriage out of 93 trials, and a probability of 0.9999 that the system would produce 64 or fewer cases of undertriage out of 93 trials. A two tailed z-test estimating the population proportions of over-and under-triage produced a Z-score of -5.1326 with a P-value of 0. The estimated population proportions are 0.312 for overtriage and 0.688 for undertriage. At the $\alpha = 0.05$ confidence level, we reject the null hypothesis that there is no difference in the likelihood of overtriage versus undertriage, when the START algorithm disagrees with the ISS-based triage category.

In Figure 14 the rates of agreement, overtriage, and undertriage relative to the ISS-based category are displayed in a clustered bar graph for the START algorithm. The vertical axis represents the numerical count of instances of agreement, over- and under- triage. The blue bar represents agreement, the red bar, overtriage, and the green bar shows undertriage. Of note, combining the cases of over-and under- triage represents greater than 50% of trials. Additionally, we see a dramatically greater preponderance of undertriage, representing the START algorithm underestimating the acuity of a given patient.

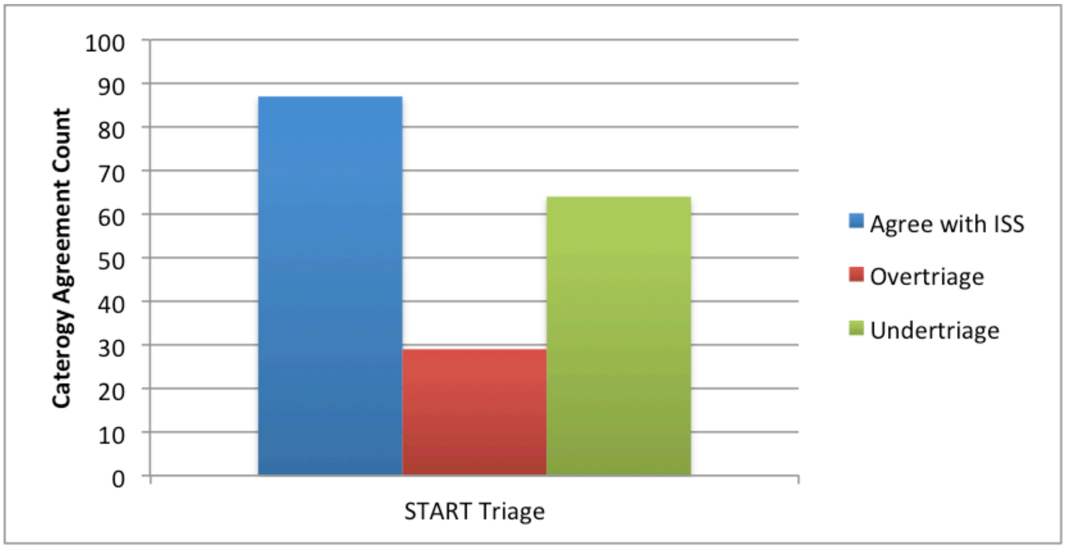

Figure 14: START Over- and Under-Triage

SALT triage. The results of the SALT algorithm agreed with the ISS-based triage category 121 times out of 180 trials. Of the 59 disagreements, 17 were overtriage, representing 28.81% of disagreements, and 42 were undertriage, representing 71.18% of cases. A binomial probability distribution resulted in a probability of 0.0007 that the system would produce 17 or fewer cases of overtriage out of 59 trials, and a probability of 0.9997 that the system would produce 42 or fewer cases of undertriage out of 59 trials. A two tailed z-test estimating the population proportions of over-and under-triage produced a Z-score of -4.6029 with a P-value of 0. The estimated population proportions are 0.288 for overtriage and 0.712 for undertriage. At the $\alpha = 0.05$ confidence level, we reject the null hypothesis that there is no difference in the likelihood of overtriage versus undertriage, when the SALT algorithm disagrees with the ISS-based triage category.

Figure 15 displays the rates of agreement, overtriage, and undertriage relative to the ISS-based category for SALT triage in a clustered bar graph. The vertical axis represents the numerical count of instances of agreement, over- and under- triage. The blue bar represents agreement, the red bar, overtriage, and the green bar shows undertriage.. As with START, we see a dramatically greater number of undertriage cases with the SALT algorithm.

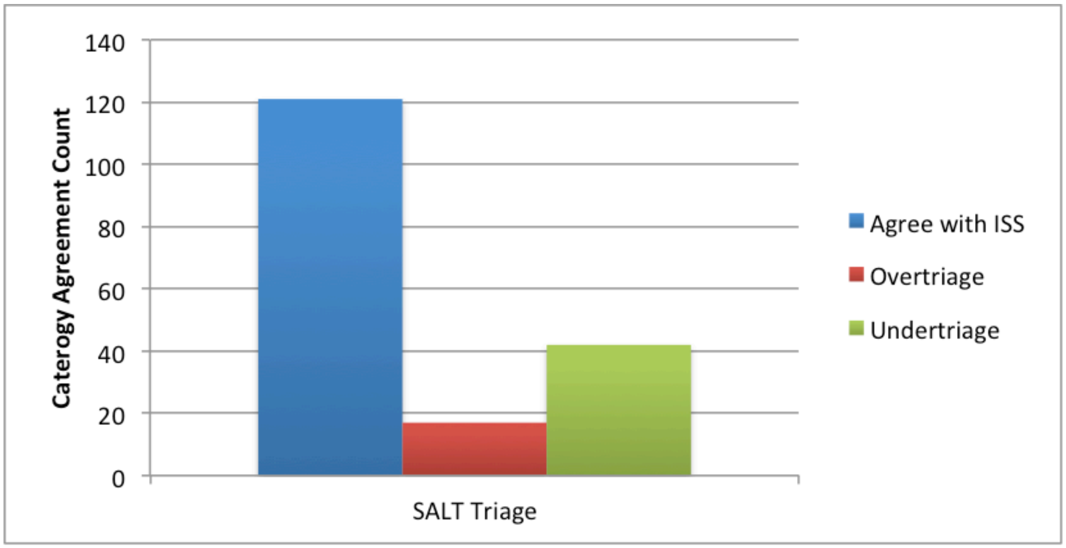

Figure 15: SALT Over- and Under-Triage

ESI triage. The results of the ESI algorithm agreed with the ISS-based triage category 66 times out of 135 trials. Of the 69 disagreements, 34 were overtriage, representing 49.27% of disagreements, and 35 were undertriage, representing 50.72% of cases. A binomial probability distribution resulted in a probability of 0.5 that the system would produce 34 or fewer cases of overtriage out of 69 trials, and a probability of 0.5950 that the system would produce 35 or fewer cases of undertriage out of 69 trials. A two tailed z-test estimating the population proportions of over-and under-triage produced a Z-score of -0.1703 with a P-value of 0.8650. The estimated population proportions are 0.493 for overtriage and 0.507 for undertriage. At the $\alpha = 0.05$ confidence level, we do not reject the null hypothesis that there is no difference in the likelihood of overtriage versus undertriage, when the ESI algorithm disagrees with the ISS-based triage category.

Figure 16 below displays the rates of agreement, overtriage, and undertriage for the ESI algorithm, in a clustered bar graph. The vertical axis represents the numerical count of instances of agreement, over- and under- triage. The bar coloration is the same as the two previous charts. Unlike START and SALT, the rates of over- and under- triage for ESI are virtually identical.

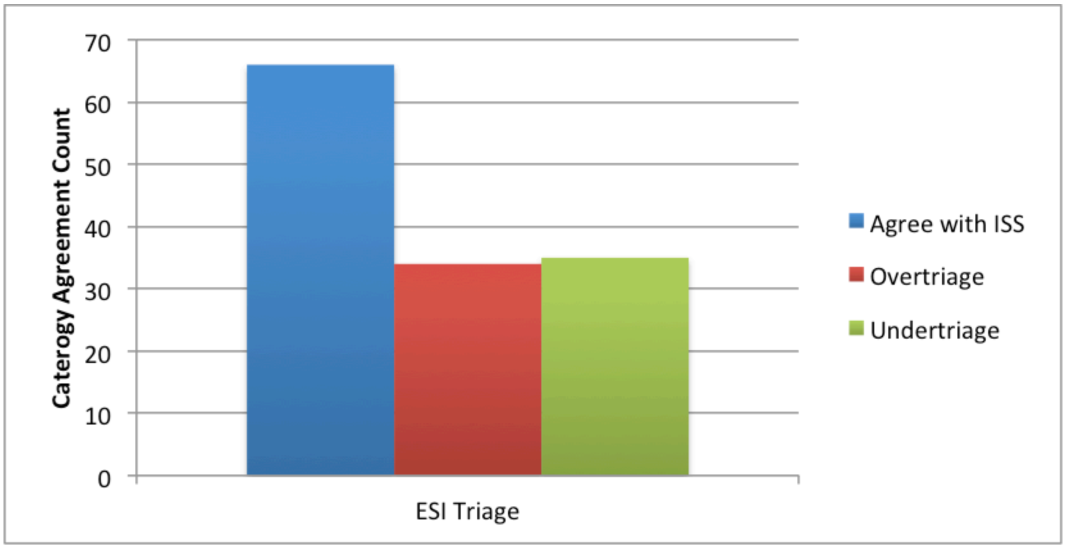

Figure 16: ESI Over- and Under-Triage

Summation of results for hypothesis four. Examining the results of testing this hypothesis, we find that in the case of START and SALT, we reject the null hypothesis that there is no difference in the rates of overtriage and undertriage for these two systems. Both tests showed a very strong tendency toward undertriage in cases when they did not produce an agreement with the ISS, as shown by their binomial probability distributions. The P-values for both *z*-tests were zero, indicating essentially no chance of a Type I error. For ESI, however, we do not reject the null hypothesis, with a P-value of 0.8650. This high P-value, as we have seen earlier, indicates there is little chance of a Type II error. The binomial proportion distributions indicate that there is essentially an equal chance that a disagreement with the ISS-based category will result in either over- or under- triage.

Discussion

To truly understand the potential value of the present work, one must consider its results in the greater context of the current state of Mass Casualty Incident Management. Currently

MCI triage exists in much the same state that Incident Management did, prior to the implementation of NIMS and the ICS structure. There is currently no clearly defined “best practice” for the performance of triage in mass casualty situation. At first glance, the MUCC (Federal Interagency Committee on EMS, 2014) would seem to be the way forward to correct this issue, but a deeper examination of the MUCC itself, and how it came about, reveal concerning issues with its widespread adoption. The MUCC was initially proposed as a concept to be developed by Lerner, et al. (2008), and was created by a team at the Centers for Disease Control and Prevention beginning in 2009. It was formally recommended by FICEMS in 2014 for widespread adoption. Even a cursory examination of the MUCC shows that it was designed explicitly to conform to the SALT algorithm, itself a product of Lerner, et al. (2008, 2011). The MUCC also explicitly excludes systems like STM which have made bold, but not yet independently confirmed claims of stronger correlation to patient mortality than even the Injury Severity Score (Sacco, et al., 2005). Of greatest concern to this writer, the MUCC was recommended for widespread adoption, with an acknowledgement that SALT was the only existing triage system that was fully MUCC compliant, a full *three years* before the publication of the first attempt to empirically validate the SALT algorithm against a reference standard (Silvestri, et al., 2017). Add to the fact that the reference standard employed in Silvestri, et al. was also the product of Lerner, et al. (2015), and itself not yet validated, and we can see that there is a grave possibility of adopting a triage methodology on a National scale that is potentially empirically provable to be less accurate at estimating patient acuity than other systems. Conducting triage in a MCI literally places the decision of “who lives” and “who dies” in the hands of the person conducting the triage; there is simply no excuse for allowing a bureaucratic decision to direct agencies to the wrong system.

In light of this context, discussion of the results is presented along two distinct tracks. First, an analysis of the data found for each hypothesis is made, looking at the data for each tested system in the light of the other two. Then an interpretation of the data combines the results of all four hypotheses tests for each triage system into a complete picture of how those hypotheses relate to each other regarding the efficacy of the discussed system. Lastly, this section includes a comparison of the results of this study to the results of preceding studies, in an attempt to relate this data to the larger body of knowledge.

Analysis of Results

Hypothesis one. The first hypothesis evaluated the tested systems conformity to the expected distribution of triage category assignment, and was as follows:

$H1_0$: Triage categories produced by the tested algorithm are consistent with the distribution of ISS-based triage categories in the sample.

$H1_a$: Triage categories produced by the tested algorithm are not consistent with the distribution of ISS-based triage categories in the sample.

Comparing the results for all three triage systems to the expected distribution of patient triage categories, we find that we must reject the null hypothesis for all three systems. In the case of this hypothesis, rejecting the null means that the observed distribution of triage results did not conform to the expected distribution; this implies that the overall product of the system is not accurate to the reference standard. Of the systems tested, the only one that came close to conformity was SALT. Consulting the x^2 distribution table, the critical value for $\alpha = 0.05$ with three degrees of freedom is 7.81. The x^2 for the SALT algorithm, 10.44, would have been significant at $\alpha = 0.01$, but such a test would have required a minimum sample size of 274, with a comparable distribution to the sample that was tested in this study. The source of this deviation from the expected distribution could be a flaw in the development of the sample; perhaps the assumptions made by this research were incorrect, and thus provided faulty information for processing the algorithm. This is discussed further in the overall interpretation of the results for the SALT algorithm. Figure 17 presents the results of all three tests of H1 in a tabular format.

Hypothesis One: Goodness of Fit			
	START	SALT	ESI
x^2 value	18.80	10.44	19.20
P-value	< 0.001	< 0.015	< 0.001
Reject $H1_0$?	Reject	Reject	Reject

Figure 17: Hypothesis One Summary

Hypothesis two. This hypothesis directly evaluated correlation between the results of the tested systems and the ISS-based triage category and the raw ISS. Stated again:

$H2_0$: There is no statistically significant correlation between the triage category assigned to the patient by the tested system and the triage category indicated by the ISS.

$H2_a$: There is a statistically significant correlation between the triage category assigned to the patient by the tested system and the triage category indicated by the ISS.

In this case the null hypothesis was rejected for all three tested systems, and with an extraordinarily high degree of certainty. While testing was conducted at the $\alpha = 0.05$ confidence level, the results would have been statistically even at $\alpha = 0.0001$. Due to the exceedingly low P-values of these test, there is a very strong likelihood that the results of the Kendall *tau* correlation accurately reflect the ability of each tested system to estimate a given patient's acuity. ESI only weakly correlated with the ISS-based triage category, however. START, the longest used formal triage system, displayed a moderate correlation. SALT, the only system fully compliant with the recommended Model Uniform Core Criteria, showed a strong correlation to the ISS-based category. All three systems showed a lesser correlation when compared directly against the Injury Severity Score itself, however. Of the three systems, ESI showed the smallest variance between the two correlation rates, only 0.0109. If the correlation scores were treated as sets of proportions and a z-test were conducted on them, none of the differences would be considered statistically significant, not even SALT, which displayed the greatest variance between correlations relative to the ISS-based triage category and the raw ISS. Comparison of SALT directly to the ISS showed only a moderate, although nearly strong, correlation. Figure 18 summarizes the results for Hypothesis Two.

Hypothesis Two: Kendall *tau* Correlation			
	START	SALT	ESI
ISS-Based Triage Category			
t-value	0.5742	0.7430	0.3874
P-value	0	0	0
Reject $H2_0$?	Reject	Reject	Reject
Raw ISS			
t-value	0.5563	0.6870	0.3765
P-value	0	0	0
Reject $H2_0$?	Reject	Reject	Reject

Figure 18: Hypothesis Two Summary

Hypothesis three. The third hypothesis set compared the rates of agreement with the ISS-based category among the three tested systems. The H3 hypothesis was:

$H3_0$: There is no statistically significant difference in the proportion of triage assignments that agree with the ISS-based category.

$H3_a$: There is a statistically significant level in the proportion of triage assignments that agree with the ISS-based category.

SALT was found to have a statistically significant difference from both START and ESI, while those two systems were not found to have a statistically significant difference from each other. This was surprising given the difference between the correlation scores between the two. If the ISS is used as a reference standard, the ESI is no more accurate than the legacy system, and both were markedly less accurate than SALT. Testing for this hypothesis also had very low P-values for the comparison of START vs. SALT and ESI vs. SALT, indicating that it is highly likely that these comparisons accurately describe the relationship between those two pairings of triage methods. The comparison of START and ESI had a very *high* P-value, which indicates that there is a strong probability that the null hypothesis is actually true. The results for Hypothesis Three are collectively shown in Figure 19, below.

Hypothesis Three: Rates of Agreement			
	START vs. SALT	START vs. ESI	ESI vs. SALT
Agreement Proportions	0.483 / 0.672	0.483 / 0.489	0.489 / 0.672
z-score	-3.6281	-0.0976	3.2785
P-value	0.00028	0.92034	0.00104
Reject $H3_0$?	Reject	Do not reject	Reject

Figure 19: Hypothesis Three Summary

Hypothesis four. The fourth hypothesis aimed to evaluate whether, when any of the tested systems were not in agreement with the reference standard, were they more likely to overestimate or underestimate the severity of a patient's injuries. The tested hypothesis was:

$H4_0$: There is no difference in the rate of under- and over- triage, relative to the category indicated by the ISS.

$H4_a$: There is a statistically significant difference in the rate of under- and over- triage, relative to the ISS category.

This speaks to the validity of the questions asked in the algorithm and the relative importance of the answers to those questions toward patient survivability. Both SALT and START showed a strong tendency to undertriage any given patient when the assigned category did not match the reference standard; nearly three of four cases of disagreement produced this result. ESI, conversely, did not have a statistically significant variance between undertriage and overtriage. This information is of practical significance, as it implies that if a responder has doubt as to the relative severity of any given patient's injuries relative to the triage category the algorithm places them into, with START and SALT, it is a fairly safe assumption to elevate the patient by one category, relative to injury severity, not treatment priority. With ESI, there is no prevailing trend that can be relied upon when the responder has an intuitive belief that the algorithm has miscategorized the patient. Testing this hypothesis also produced very small P-values for the START and SALT algorithms, and a high P-value for ESI. This shows that the decision to reject (in the case of START and SALT) or not reject (in the case of ESI) is very likely correct. Hypothesis Four's results are shown in Figure 20.

Hypothesis Four: Rates of Over- and Under- Triage			
	START	SALT	ESI
Overtriage Proportion	0.3118	0.2881	0.4927
Overtriage Binomial Probability	0.0001	0.0007	0.5
Undertriage Proportion	0.7419	0.7118	0.5072
Undertriage Binomial Probability	0.9999	0.9997	0.5950
z-score	-5.1326	-4.6029	-0.1703
P-value	0	0	0.8650
Reject $H4_0$?	Reject	Reject	Do not reject

Figure 20: Hypothesis Four Summary

Interpretation and Conclusions

The tested hypotheses each reveal some small portion of a larger picture regarding the efficacy of the three subject triage algorithms. Looking at each of the results in the context of the others allows us the opportunity to draw conclusions about these systems, as well as about other research that has been previously conducted. Regarding START and SALT in particular, there is value in comparing the results of this study to the work of Silvestri, et al. (2017) and others.

START triage. The START triage system represents the "Grand Old Man" of mass casualty incident management. Until very recently, it was also the only system in use in the United States that had been empirically tested (Hong, et al., 2007). It is also the most widely employed triage system in the United States. As such, it can be thought of as a practical 'gold standard', even if it cannot be thought of as such from an empirical perspective. In order to justify the expense, time, effort, and risk of poor patient outcomes that would result from any agency's decision to transition from START to another triage method, the evidence supporting the proposed change must be highly compelling. In this study, we find first that the assignment of triage categories did not conform to the expected distribution. As noted above, this could potentially speak to error on the part of the researcher. If the data in Figure 4 is compared to the data shown in Figure 13, however, one may reasonably draw the conclusion that this deviation reflects an error in the algorithm itself, based on the nearly 3 to 1 ratio of undertriage to overtriage, and that the preponderance of deviation from the expected distribution lies in the minimal category. Looking at the algorithm itself, this becomes even more clear; adhering strictly to the algorithmic decision tree, a patient suffering from an open pneumothorax (universally considered to be an "Immediate" category injury), but who retained the ability to walk, would be miscategorized by two levels as a minimal patient. Only the clinical experience of the rescuer would prevent this error.

START does show a moderate degree of correlation to both the ISS-based triage category and the raw ISS. It did, however, arrive at the same triage assignment in less than 50% of cases, and 74% of the errors were found to underestimate the severity of a patient's injury. Put in clearer terms, this means that 36 times out of 100, the START algorithm produces a result that indicates a patient is less severely injured than they actually are. In the mind of this researcher, this represents an unacceptable level of risk.

SALT triage. If START is the "Grand Old Man" of triage systems, SALT is the young upstart. Currently only a decade old, it lacks historical validation, but it did have the benefit of an additional 25 years of research to draw from in its development. It also currently has the benefit of strong backing with the endorsement of the MUCC by FICEMS (while not a direct endorsement of SALT, it can be viewed as one since SALT is the only triage system that currently complies with all 24 MUCC recommendations), as well as the selection by the National Disaster Life Support Foundation as their triage method of choice.

Like START, SALT did not conform to the expected distribution of triage assignments. Comparing Figures 5 and 14, we see that the preponderance of deviation lies in the Immediate, Delayed, and Expectant categories. This does represent a greater likelihood of researcher error; specifically the assumption made that any patient who arrived at the ED with signs of life was "likely to survive given the current resources". Looking at the raw data, this represents a total of four cases. If this were a researcher error, all four of these cases would be miscategorized from Expectant to Immediate. Adjusting the category counts to reflect this error, we would find that SALT produced 46 Minimal, 54 Delayed, 49 Immediate, and 31 Expectant. Recalculating the Goodness of Fit test for these counts and $\alpha = 0.05$ yields $x^2 = 6.533$ and a P-value of 0.088, indicating that the results are *not* statistically significant, the null hypothesis should *not* be rejected, and that START *does* in fact conform to the expected distribution. Figure 21 displays the corrected Goodness of Fit data below, and is organized identically to Figure 6 above.

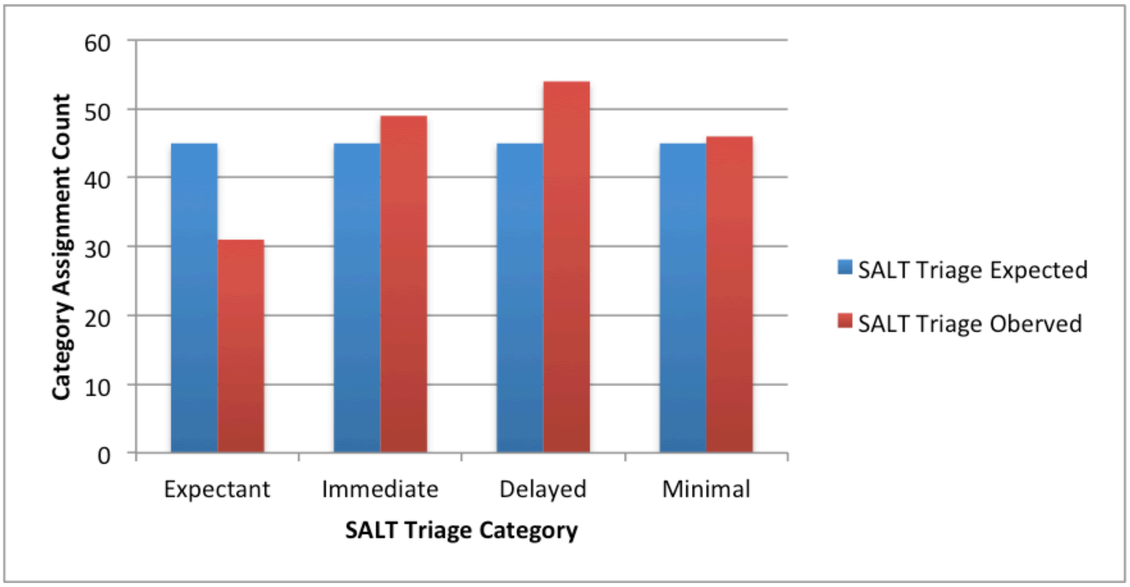

Figure 21: Corrected SALT Triage Goodness of Fit

Recalculating the Kendall *tau* with the adjusted results increases the correlation coefficient to 0.7821 versus the ISS-based category, and 0.7221 against the raw Injury Severity Score, both with a P-value of 0.0000. This adjustment increases the correlation strength to both the ISS-based category and the raw ISS to ‘strong’ for both, versus ‘strong’ and ‘moderate. Figure 22 presents the correlational data in the same manner as Figure 9.

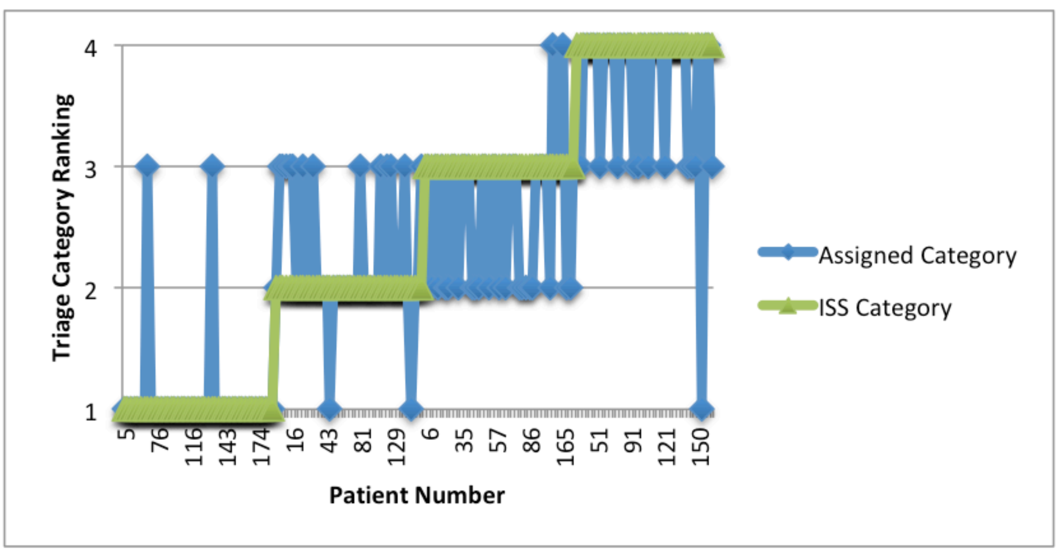

Figure 22: Corrected SALT Correlation

Adjusting the patient counts to correct the rate of agreement for SALT produces 125 cases of agreement out of 180 trials, or 69.44% of cases. Conducting a two-tailed z-test on the difference in proportions between START versus. SALT, and ESI versus SALT, produces Z-scores of -4.0704 at a P-value of 0 and -3.6954 at a P-value of 0.0022 respectively. Again we see the SALT algorithm increase its strength against the other two tested systems. Figures 23 and 24 show the resulting comparisons in the same manner as Figures 11 and 13, respectively.

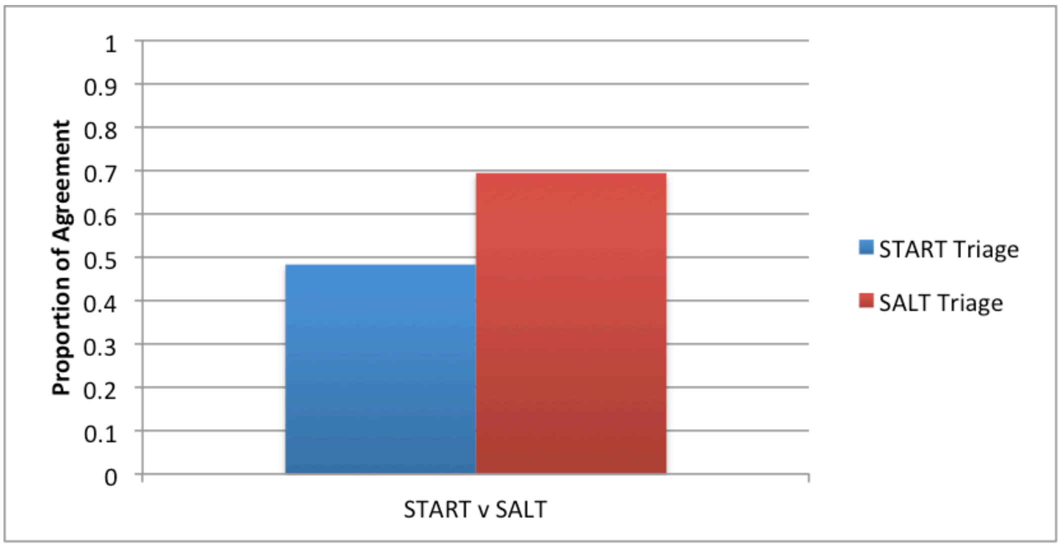

Figure 23: Corrected START versus SALT Proportions of Agreement

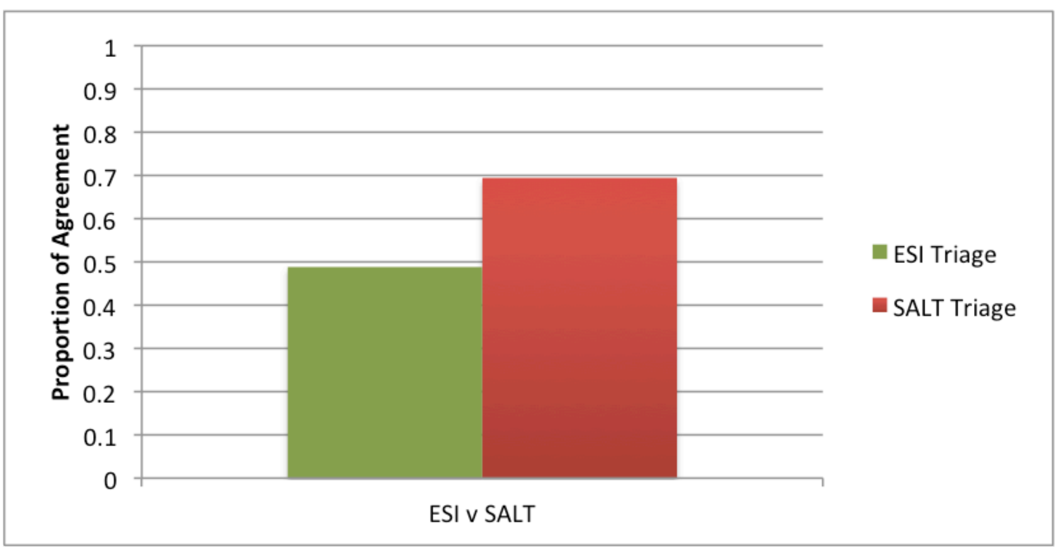

Figure 24: Corrected ESI versus SALT Proportions of Agreement

Lastly, evaluating the rates of over- and under-triage, we find that overtriage represents 17 of 55 (30.90%) cases and undertriage represents 38 of 55 (69.09%) cases of disagreement with the ISS-based triage category. Recalculating the binomial probability of over- and under-triage with the adjusted scores, we find that there is a 0.0032 probability that the SALT algorithm will produce 17 or fewer cases of overtriage and a 0.9987 probability of 38 or fewer cases of undertriage, when it does not agree with the ISS-based category. A two-tailed z-test based on these adjustments yield a Z-score of -4.0045 and a P-value of 0. At $\alpha = 0.05$, we would still reject the null hypothesis. Assuming that there is researcher error only strengthens the results of each tested hypothesis, rather than weakening or invalidating them. Figure 25 displays the corrected over- and under- triage counts, corresponding to the chart structure of Figure 15, above.

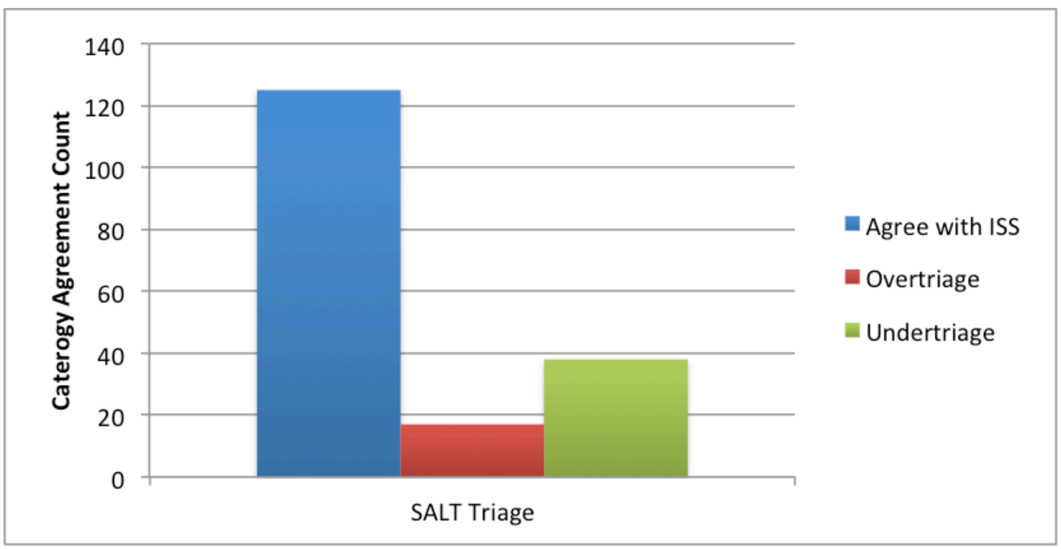

Figure 25: Corrected SALT Over- and Under-Triage

Looking at the results of SALT in the same light as we did START, we would find that SALT strongly correlates to the ISS-based triage category, and moderately to the raw ISS. Adjusting for researcher error increases the correlation with the raw ISS to a strong one. Additionally, SALT agreed with the ISS-based category in either 67.22% or 69.44% of cases, and the rate of undertriage in cases of disagreement was found to be either 71.18% (assuming no researcher error), or 69.09% (assuming researcher error). This means that the number of patients that SALT would consider to be more severely injured than they really were is 23 out of 100, or 21 out of 100, depending on whether the researcher error exists or not. This represents a 36.11% or 41.67% reduction in the number of undertriaged patients, relative to the START algorithm. Figure 26 provides all of the corrected data points for SALT Triage in a single table.

Summary of Corrected Results for SALT Triage

Hypothesis One

x^2 value	P-value	Reject $H1_0$?
6.533	0.088	Do not reject

Hypothesis Two

ISS-Based Triage Category			Raw ISS		
t-value	P-value	Reject $H2_0$?	*t*-value	P-value	Reject $H2_0$?
0.7821	0	Reject	0.7221	0	Reject

Hypothesis Three

START vs. SALT				ESI vs. SALT			
Agreement Proportion	*z*-value	P-value	Reject $H3_0$?	Agreement Proportion	*z*-value	P-value	Reject $H3_0$?
0.483 / 0.694	-4.0704	0	Reject	0.489 / 0.694	-3.6954	0.0022	Reject

Hypothesis Four

Overtriage Proportion	Overtriage Binomial Probability	Undertriage Proportion	Undertriage Binomial Probability	*z*-value	P-value	Reject $H4_0$?
0.3090	0.0032	0.6909	0.9987	-4.0045	0	Reject

Figure 26: Summary of Corrected Data SALT Triage

ESI triage. ESI was developed specifically for use in the Emergency Department, in an attempt to more effectively refine the order in which patients received treatment. Notably the intent of the system was to increase the number of patient categories from two or three, to five. (Wuerz, et al., 2000). While this aim was accomplished, it is worth noting that the resulting system makes no accommodation for patients who are expired or unlikely to survive. This is likely due to a misconception that EDs do not have resource constraints. The problem with this line of thinking is that in an event such as the September 11, 2001 attack on the World Trade Center, an individual ED may be completely overrun with casualties, essentially moving the mass casualty incident from the point of injury to the parking lot or waiting room. ESI, qualitatively, is wholly unsuitable for use on the scene of a mass casualty incident, specifically because of the omission of an Expectant or Dead category. If a hospital employs ESI for their routine triage, they will need to employ a different system in a mass casualty incident. This presents an increased likelihood of inappropriate triage decisions due to lack of familiarity with the triage method being used.

Regarding the possibility of researcher error in developing the triage handouts for the ESI algorithm, the most likely point of error would be at the decision point regarding danger zone vitals. The decision to direct the volunteers to elevate any patient from Category 3 (Minimal) to Category 2 (Delayed) stemmed from the attempt to avoid asking the volunteers to make a judgment call. The ESI algorithm states that the patients in question should be considered for elevation to Category 2. In the absence of an actual patient to evaluate in totality, the decision was made to always elevate these patients to the next higher triage category. If this decision were a source of error in the results, we would expect to find a significantly higher number of patients overtriaged to Category 2 than undertriaged to Category 3. This was found to be true; there were 22 of 45 patients overtriaged to Delayed versus 11 of 45 patients undertriaged to Minimal. Running a two-tailed z-test on these two proportions resulted in a Z-score of 2.4061 with a P-value of 0.01596. At $\alpha = 0.05$, this potentially represents a statistically significant error. Given that this decision point is entirely up to the judgment of the person performing triage, we then compared the total rates of disagreement between the ESI algorithm and the ISS based category for each triage category independent of the others. This showed 26 of 45 cases of disagreement in the Minimal category, 19 of 45 cases in the Delayed category, and 24 of 45 cases in the Immediate category. Each of these proportions were compared to the other two in a two tailed z-test for statistical significance, and none of the pairings were found to be significant at $\alpha = 0.05$. In light of this result, no correction to account for researcher error is deemed necessary regarding the ESI algorithm.

In this study, ESI proved, in the case of every hypothesis, to be inferior to both the more recently developed SALT algorithm and the legacy START system. Comparing the lower correlation coefficients found against the raw ISS, ESI correlated 32% less strongly to the reference than did START, and at least 45% less strongly than SALT. If the assumption is made that this researcher erred in the development of the test data for the SALT algorithm, that number increases to 48%. The only area that can be found in which ESI was not wholly inferior to the other two tested systems, was in the difference between its correlation to the ISS-based category and its correlation to the raw ISS. That difference was only 0.0109, but given the overall weakness of the correlation, this cannot be considered a saving grace.

Conclusions. The data presents a clear and compelling picture regarding triage methodology. Of the three tested systems, in every regard, SALT is the superior method of evaluating patient acuity. START is the next most effective, but significantly lacking in both accuracy and the rate of undertriage. ESI is the least effective of the evaluated systems; it only weakly correlates to the Injury Severity Score, which has a well established relationship to patient mortality (Bull, 1975). In most cases, the ESI methodology does not appropriately categorize patients according to their acuity, and this is the system that is intended for use at definitive care facilities.

Comparison of results to preceding studies.

When we compare the results of this present work to some of the studies that have preceded it, we find the present work to add definitive value to the field. Without over-interpreting the data presented above, we find that the present work appears to validate the results of at least three prior works. Perhaps the most immediate, and intuitively satisfying, comparison between this work and preceding works is to that of Silvestri, et al. (2017). In their research, they evaluated rates of agreement to the reference standard developed by Lerner, et al. (2015) for both the START and SALT triage systems. Their results indicated 49 of 82 cases of agreement for START and 61 of 82 agreements using the SALT system. This result provides an excellent comparison point to the present study, matching precisely to the methodology employed in the H3 hypothesis. Comparing these rates with a two-tailed z-test provided the following results: START yielded a Z-score of 1.716 and P-value of 0.08544, and SALT resulted in a Z-score of 1.1682 and P-value of 0.242. Correcting for researcher error as described above, the results for SALT change to Z-score = -0.2025 and P-value of 0.84148. At $\alpha = 0.05$, none of these differences are statistically significant, and the direct proportions of agreement, especially when using the corrected SALT results, are virtually identical. This would imply not only the validity of Silvestri, et al. (2017), but also the validity of Lerner, et al. (2015). While direct testing would still be of value in the case of Lerner, et al., the implication is that their consensus based 'gold standard' for triage category assignment would correlate well to the Injury Severity Score, and thus to patient mortality rates.

Comparing the current work to that of Kosashvili, et al. (2009), also presents interesting and valuable insights into the management of mass casualty incidents. Their observational study of bombing mass casualty incidents in Israel was the direct source for the ISS bands used in the current work. In all three systems, the correlation to the ISS-based category was stronger than to the raw ISS score. This could be explored further to determine the exact ISS bands that correlate most strongly with the five tiered, color coded system recommended by the MUCC. Since Kosashvili et al. focused specifically on hospital surge capacity, and their study found that consistently 20% of bomb related mass casualty incident victims required hospitalization, there is an implied correlation between the total number of victims of a mass casualty incident and an expected distribution of triage category assignments. Further examination of this potential correlation could provide useful data in the management of these types of events, particularly in the area of personnel and material resource management.

Recommendations

There are a number of recommendations that can be drawn from the results of this study, ranging from recommendations for additional academic examination, to modifications to the development of exercise scenarios, to very practical, street level changes to agency policies. Some of those recommendations are listed below.

Recommendation I

A correlational study relating the ISS to the consensus arrived at by Lerner, et al. (2015) should be conducted. Showing that their results are positively correlated to the ISS would provide backing for additional research into triage methodologies using the ISS as a proxy for their standards. This is important because the consensus developed by Lerner, et al. can only be used to retrospectively evaluate a patient record and determine the triage category they *should have* been assigned. The Injury Severity Score, while not directly useful to field triage, could prove valuable in a definitive care setting, as it can be determined once all of the ICD-10 diagnostic codes for a given patient have been identified. This could potentially be used as a secondary check on the field triage decision, and could help to identify patients that had been undertriaged in the field.

Recommendation II

Studies are needed to determine the exact ISS transition points between color based triage categories, and to what degree those categories overlap. This could be done once the correlation between Lerner, et al. (2015) and the ISS was established, by retrospectively classifying patient records using Lerner, et al.'s standard and comparing it to the ISS score for that patient. In particular, identifying the areas of overlap could help to refine triage algorithms by focusing on finding points of distinction within those overlaps.

Recommendation III

While the SALT algorithm has emerged as a clear practical 'gold standard' for triage, being the most effective system currently in common usage, this does not mean that it is the most accurate method of triaging casualties. In particular, the Sacco Triage Method (STM) shows great promise, although it has had limited independent evaluation, and this researcher is aware of no study that evaluates its efficacy against the SALT algorithm. Such a study is sorely needed, as the STM methodology was presented by Sacco, et al. (2005) as superior to the ISS as a predictor of patient mortality. Given an implied correlation between ISS and Lerner, et al. (2015) and the demonstrated correlation between ISS and the SALT method, this could mean that STM is more effective both as a triage system and a means of evaluating novel systems. This comparison would also have value to the MUCC. Systems compliant with the core criteria are mandated to be five tier systems (Federal Interagency Committee on EMS, 2014). Proponents of STM have argued against (ThinkSharp, Inc., 2012), and this tiering requirement is one of the two areas in which the STM is not MUCC compliant (the other being the avoidance of use of timed-counting methods of assessing a patient, such as counting respirations or taking a pulse).

Recommendation IV

The search for the 'best' triage system should be expanded beyond national borders, and comparison studies should be conducted to evaluate the various triage systems employed globally. Establishing an international benchmark for triage system effectiveness will pay great dividends as disasters become increasingly more likely to involve a multinational response, with the potential for poor patient outcomes resulting from the application of multiple triage methods and designation patterns.

Recommendation V

Research should be conducted to evaluate the various systems in light of the ocean of data that has emerged from the last 17 years of warfare in the Middle East, which has resulted in the development and growing implementation of the Tactical Combat Casualty Care and Tactical Emergency Casualty Care protocols. This data has shown that the traditional patient management methods employed by ATLS protocols are not only less effective at the management of penetrating trauma, in some cases they are detrimental to patient survival. Given the current understanding of the vast difference in injury physiology and treatment methods between blunt and penetrating trauma, it is not unreasonable to arrive at the conclusion that triage methods might likewise need to be altered depending on the mechanism of injury. This should be thoroughly examined.

Recommendation VI

The academic community should discontinue the practice of comparing novel triage systems to START. At this point in the development of the body of knowledge, enough studies have shown the START algorithm to be lacking that continuing to use it as a reference lacks rigor, and would represent a practice of posterity only. The only legitimate value to be gained by comparing a novel system to START would be in a study evaluating the ease of application of a novel system by the personnel of a specific jurisdiction, provided that the studied jurisdiction currently employs the START algorithm.

Recommendation VII

The Emergency Response community needs to alter the way that develops and conducts exercises in light of the current science. Currently, there is a relatively prevalent attitude that triage is a somewhat subjective endeavor, and as such its performance during exercises cannot be easily evaluated objectively. This should be put to rest. There is clear and mounting empirical data that shows that any given trauma patient can be described in objective terms with regard to which triage category they should be assigned to. Additionally, as the ISS is based solely on the specific injuries a patient has sustained, an exercise designer can calculate it without needing to identify the treatments a patient would need to receive at definitive care, or speculate on the timelines to those interventions. Mass casualty exercise design should be adjusted to identify the ISS for each patient in the scenario, so that an objective grading criteria can be employed in the evaluation of responder actions.

Recommendation VIII

The after action reports of exercises developed in the manner recommended above should be treated as a data source for meta-analysis studies, as the number of mass casualty exercises far outstrips the number of actual mass casualty incidents. If triage is graded objectively, and data is collected specifically regarding the cases of disagreement from the reference standard, it would provide a tremendous resource to assist in determining the failure points of whatever triage system is employed by that agency. This currently untapped resource would be a great boon to researchers in the area of disaster medicine and mass casualty incident management.

Recommendation IX

The Model Uniform Core Criteria should be modified prior to adoption on a national scale. Even a cursory look at these criteria shows a very evident bias toward the SALT algorithm; the 24 criteria are actually grouped according to the SALT acronym. Additionally, the insistence on the use specifically of a five-tiered, color-coded designation system presupposes that such a system is inherently more clear and easier to understand than other methods. This does not account for the possibility of color blindness among the rescuer population, and likewise the assumption that a single number indicating a patient's ranking in the system is somehow confusing seems preposterous. The national implementation of the MUCC as it currently stands will inevitably drive states to mandate the use of the SALT algorithm, as it is the easiest path to regulatory compliance. This is not problematic if SALT can be shown empirically to be the most effective triage system, as it was in this study. Given that STM and the various international systems have not yet been evaluated against SALT, it seems irresponsibly premature to assume this to be the case.

Recommendation X

Lastly, and perhaps most pressing, the use of ESI as a triage method in any setting should be discontinued. The results of this study show that it is completely inferior in every evaluated aspect to SALT, and even to the dated and provably problematic START. Any agency that utilizes this system faces the real risk of poor patient outcomes and resulting litigation if they knowingly continue to employ a triage system that can be shown to be 32.7% to 52.5% less effective at identifying the most acutely injured patients, and by extension, the patients at most risk for mortality. It is worth noting that this researcher does not directly advocate for agencies to convert their triage protocols from START to the SALT system, especially in light of the need for further evaluation against a number of other triage methodologies. It is the opinion of this writer that a large-scale adoption of the SALT algorithm runs the risk of quickly running afoul of further research. Any agency that does elect to transition to the SALT system would be making a step in the right direction, reducing the risk of patient mortality, but would need to do so with the understanding that while SALT currently shows the best correlation to patient acuity, there is a substantial possibility that one of the other systems currently in use will supplant it. If that were to happen, the agency in question could end up implementing another transition in short order, if they desire to utilize the current state of the art in triage methods.

Conclusion

This study aimed to determine which, if any, of three triage methods commonly employed in the United States correlated most strongly to the well documented reference standard of the American College of Surgeons' Injury Severity Score. The three tested systems, Simple Triage and Rapid Treatment (START), Sort, Assess, Lifesaving interventions, and Triage/Treatment (SALT), and the Emergency Severity Index (ESI) were each evaluated on how well their results conformed to the expected patient distribution, their correlation to both a triage category based on Injury Severity Score bands, and to the Injury Severity Score itself. An evaluation was also made for the rate at which the three systems agreed with the ISS-based triage category, with a direct comparison made between each unique pairing of systems on this point. Finally, the tested systems were examined to determine if, in the case of a disagreement with the reference standard, they were more likely to over- or under- triage any individual patient. This point reflects the system's tendency to either overestimate or underestimate the severity of any given patient's injuries.

One hundred eighty patients were selected at random from the National Trauma Data Bank 2016 Research Data Set, and those records were utilized to generate a set of answers to each of the questions asked by the three tested systems. This allowed the research to evaluate the three tested systems in isolation from the individual clinical knowledge and experience level of the person actually conducting triage, more closely tying the results of the study to the validity of the triage systems themselves. Volunteers were provided these answers and a triage algorithm and were asked to determine which category their system would place that patient into. Each patient was evaluated once with each triage system, with the exception of the 45 patients selected with an ISS of 75; ESI does not include an expectant category, and it was decided not to evaluate these patients using that system, as they would skew the results of the study against ESI.

The resulting triage assignments were then compared to the reference category and a number of statistical tests were conducted. Analyzing and interpreting the results of these tests identified a possible error in the design of the study. Correcting for this potential error was found to only strengthen the ultimate results of the study.

The SALT algorithm was found to be superior in every tested parameter to both START and ESI, and with the correction of the potential error, was found to conform to the expected distribution of triage category assignments, exhibit a strong correlation to both the ISS-based triage category and the raw ISS, have a markedly better rate of agreement with the reference standard, and displayed a lower rate of both over- and under- triage from the other testes systems. In cases of disagreement with the reference standard, a tendency to underestimate the severity of an individual patient's injuries was still observed. Statistical testing of all four hypotheses for all three systems exhibited a high degree of surety, indicating the validity of the quantitative rejection or failure to reject the null hypotheses.

These results were noted to validate a number of preceding studies, and strengthen the claim that SALT is the most effective triage system in common usage in the United States. Further studies would be needed to cement that claim, however, as well as to develop an international triage standard. A number of recommendations can be drawn from the results of this study, and should be implemented.

Determining the most effective means of identifying the most acutely injured patients in a mass casualty incident is literally a matter of life and death. The errors inherent in the tested systems directly represent patients either not receiving the care that they require to survive or receiving care that would be better applied to other patients in the context of the full incident. Continued efforts in this regard will help to identify and eliminate the points of variation within the systems that lead to errors in triage category assignment, ultimately enhancing patient outcomes and survivability.

References

Agency for Healthcare Research and Quality. (2014, October). ESI Triage Algorithm [Digital image]. Retrieved January 20, 2018, from https://www.ahrq.gov/professionals/systems/hospital/esi/esifig21a.html

Bhalla, M. C., Frey, J., Rider, C., Nord, M., & Hegerhorst, M. (2015). Simple Triage Algorithm and Rapid Treatment and Sort, Assess, Lifesaving, Interventions, Treatment, and Transportation Mass Casualty Triage Methods for Sensitivity, Specificity, and Predictive Values. *American Journal of Emergency Medicine*, 1687-1691.

Bull, J. P. (1975). The Injury Severity Score of Road Traffic Casualties in Relation to Mortality, Time of Death, Hospital Treatment Time and Disability. *Accident Analysis & Prevention, 7*(4), 249-255.

Butler, F. K., Hagmann, J., & Butler, E. G. (1996). *Tactical Combat Casualty Care in Special Operations*.

Cone, D. C., & MacMillan, D. S. (2005). Mass-Casualty Triage Systems: A Hint Of Science. *Academic Emergency Medicine, 12*(8), 739-741. doi:10.1197/j.aem.2005.04.001

Culley, J. M., Svendsen, E., Craig, J., & Tavakoli, A. (2014). A Validation Study of 5 Triage Systems Using Data from the 2005 Graniteville, South Carolina, Chlorine Spill. *Journal of Emergency Nursing, 40*(5), 453-460.

Eastridge, B. J., M.D, Butler, F., M.D, Wade, C. E., Ph.D, Holcomb, J. B., M.D, Salinas, J., Ph.D, Champion, H. R., M.D, & Blackbourne, L. H., M.D. (2010). Field Triage Score (FTS) in Battlefield Casualties: Validation of a Novel Triage Technique in a Combat Environment. *The American Journal of Surgery, 200*(6), 724-727. 10.1016/j.amjsurg.2010.08.006

Hong, R., Sierzenski, P. R., O'Connor, R. E., Bollinger, M., & Durie, C. C. (2007). 38: Does the Simple Triage and Rapid Treatment Method Appropriately Triage Patients Based on Trauma Injury Severity Score? *Annals of Emergency Medicine, 50*(3), S13-S13. doi:10.1016/j.annemergmed.2007.06.069

Garner, A., Lee, A., Harrison, K., & Schultz, C. H. (2001). Comparative Analysis of Multiple-Casualty Incident Triage Algorithms. *Annals of Emergency Medicine, 38*(5), 541-548.

González, P. A., Delgado, R. C., Alvarez, T. C., Gonzalo, G. G., Monzon, C. M., Corres, N. P., ... & Fuentes, F. T. (2016). The Development and Features of the Spanish Prehospital Advanced Triage Method (META) for Mass Casualty Incidents. *Scandinavian Journal Of Trauma, Resuscitation and Emergency Medicine, 24*(1), 63.

Federal Interagency Committee on EMS. (2014, March). *National Implementation of the Model Uniform Core Criteria for Mass Casualty Incident Triage (Rep.)*. Retrieved March 19, 2018, from website: http://www.nhtsa.gov/staticfiles/nti/pdf/811891-Model_UCC_for_Mass_Casualty_Incident_Triage.pdf

Jenkins, J. L., McCarthy, M. L., Sauer, L. M., Green, G. B., Stuart, S., Thomas, T. L., & Hsu, E. B. (2008). Mass-Casualty Triage: Time for an Evidence-Based Approach. *Prehospital and Disaster Medicine, 23*(1), 3-8.

Khan, A., Mahadevan, S. V., Dreyfuss, A., Quinn, J., Woods, J., Somontha, K., & Strehlow, M. (2016). One-Two-Triage: Validation and Reliability of a Novel Triage System for Low-Resource Settings. *Emergency Medicine Journal: EMJ, 33*(10), 709-715.

Kilner, T. (2002). Triage Decisions of Prehospital Emergency Health Care Providers, Using a Multiple Casualty Scenario Paper Exercise. *Emergency Medicine Journal: EMJ, 19*(4), 348-353. doi:10.1136/emj.19.4.348

Kosashvili, Y., Daniel, L. A., Peleg, K., Horowitz, A., Laor, D., & Blumenfeld, A. (2009). Israeli Hospital Preparedness for Terrorism-Related Multiple Casualty Incidents: Can the Surge Capacity and Injury Severity Distribution be Better Predicted? *Injury, 40*(7), 727-731. doi:10.1016/j.injury.2008.11.010

Lerner, E. B., Schwartz, R. B., Coule, P. L., Weinstein, E. S., Cone, D. C., Hunt, R. C., . . . O'Connor, R. E. (2008). Mass Casualty Triage: An Evaluation of the Data and Development of a Proposed National Guideline. *Disaster Medicine and Public Health Preparedness, 2 Suppl 1*(Supplement 1), S25-S34. doi:10.1097/DMP.0b013e318182194e

Lerner, E. B., Cone, D. C., Weinstein, E. S., Schwartz, R. B., Coule, P. L., Cronin, M., . . . Hunt, R. C. (2011). Mass Casualty Triage: An Evaluation of the Science and Refinement of a National Guideline. *Disaster Medicine and Public Health Preparedness, 5*(2), 129-137. doi:10.1001/dmp.2011.39

Lerner, E. B., McKee, C. H., Cady, C. E., Cone, D. C., Colella, M. R., Cooper, A., … Swienton, R. E. (2015). A Consensus-Based Gold Standard for the Evaluation of Mass Casualty Triage Systems. *Prehospital Emergency Care : Official Journal of the National Association of EMS Physicians and the National Association of State EMS Directors, 19*(2), 267–271. http://doi.org/10.3109/10903127.2014.959222

Model Uniform Core Criteria for Mass Casualty Triage. (2011). *Disaster Medicine and Public Health Preparedness, 5*(2), 125-128. doi:10.1001/dmp.2011.41

Sacco, W. J., Navin, D. M., Fiedler, K. E., Waddell II, R. K., Long, W. B., & Buckman, R. F. (2005). Precise Formulation and Evidence-Based Application of Resource-Constrained Triage. *Academic Emergency Medicine, 12*(8), 759-770. doi:10.1197/j.aem.2005.04.003

Silvestri, S., Field, A., Mangalat, N., Weatherford, T., Hunter, C., McGowan, Z., ... & Ralls, G. (2017). Comparison of START and SALT Triage Methodologies to Reference Standard Definitions and to a Field Mass Casualty Simulation. *American Journal of Disaster Medicine, 12*(1), 27-33.

Stangroom, J. (2018). *Social Science Statistics*. Retrieved March 16, 2018, from http://www.socscistatistics.com/Default.aspx

START Adult Triage [Digital image]. (2017, September 29). Retrieved January 21, 2018, from https://chemm.nlm.nih.gov/chemmimages/StartAdultTriageAlgorithm.gif

[SALT Triage Algorithm]. (2017, September 29). Retrieved January 21, 2018, from https://chemm.nlm.nih.gov/chemmimages/salt.png

Streckbein, S., Kohlmann, T., Luxen, J., Birkholz, T., & Prückner, S. (2016). Triage Protocols for Mass Casualty Incidents : An Overview 30 Years After START. *Der Unfallchirurg, 119*(8), 620.

ThinkSharp, Inc. (2012). *STM & National Triage Initiatives*. Retrieved March 23, 2018, from http://saccotriage.com/stm-national-triage-initiatives.html

US Department of Health and Human Services. (2017, September 29). *START Adult Triage Algorithm - CHEMM*. Retrieved January 15, 2018, from https://chemm.nlm.nih.gov/startadult.htm

Vasavada, N. (2016). *Online Web Statistical Calculators*. Retrieved March 16, 2018, from http://astatsa.com/

Wuerz, R. C., Milne, L. W., Eitel, D. R., Travers, D., & Gilboy, N. (2000). Reliability and validity of a new Five‐level triage instrument. *Academic Emergency Medicine, 7*(3), 236-242. 10.1111/j.1553-2712.2000.tb01066.x

Wuerz, R. C., Travers, D., Gilboy, N., Eitel, D. R., Rosenau, A., & Yazhari, R. (2001). Implementation and Refinement of the Emergency Severity Index. *Academic Emergency Medicine, 8*(2), 170-176. doi:10.1111/j.1553-2712.2001.tb01283.

Appendix A

Thank you for choosing to help me complete this study project. This study will be assessing the accuracy of three different triage systems at placing individual patients into the correct triage category, based on the reference of the American College of Surgeons' Injury Severity Score. Your assistance will prove invaluable in eliminating the possibility of bias from my study. In order to ensure that my study is not considered to be research on human subjects, with all the regulatory concerns that entails, I will not be collecting any information about the volunteers who choose to participate. You will not be individually recognized in the final paper, but your efforts will be acknowledged collectively.

START Triage Instructions

You have been provided with a copy of the START triage algorithm, these instructions, and a list of information about patients. The list of patients will provide you only with the specific information to answer the questions asked by the triage algorithm. Your task is to follow the algorithm until reaching a category assignment, and note the category in the block provided for each patient. Every patient entry will contain the answers to ALL of the questions asked by the algorithm, not only the ones needed to assign the triage category. Once you have arrived at a category assignment, disregard any further information in that patient's entry, and do not continue further on the algorithm with that patient. Simply annotate the first triage category that the algorithm leads you to.

There is only one deviation from the standard algorithm for you to be aware of:

-The source data used to generate the patient records does not provide information on whether repositioning the airway was successful at restoring spontaneous respirations. Instead, the presence of spontaneous respirations is grouped with successful airway repositioning of the airway.

Patient number	Ambulatory?	Spontaneous Breathing / airway repositioning successful?	Respiratory Rate above or below 30	Radial Pulse Present / Cap refill < 2 sec?	Obeys Commands?
1	y	y	Below	y	y
2	n	y	Below	y	y
3	y	y	Above	y	y
4	n	n	Below	y	n
5	y	y	Below	y	y
6	n	y	Below	y	n
7	y	y	Below	y	y
8	y	y	Below	y	y
9	y	y	Below	y	y
10	n	y	Below	n	y
11	n	n	Below	y	n
12	y	y	Below	y	y
13	n	y	Below	y	n
14	n	y	Below	y	y
15	n	y	Below	y	n
16	y	y	Below	y	y
17	y	y	Below	n	y
18	n	y	Below	y	y
19	y	y	Below	y	y
20	y	y	Below	y	y
21	y	y	Below	y	y
22	y	y	Below	y	y
23	y	y	Below	y	y
24	n	y	Below	y	y
25	y	y	Below	y	y
26	n	y	Below	y	n
27	n	n	Below	y	n
28	n	y	Below	y	y
29	n	y	Below	n	n
30	n	n	Below	y	n
31	n	y	Below	y	y
32	n	y	Below	y	y
33	n	n	Below	n	n
34	n	n	Below	n	n
35	n	y	Below	y	n
36	n	y	Below	y	n
37	n	y	Below	y	y

Patient number	Ambulatory?	Spontaneous Breathing / airway repositioning successful?	Respiratory Rate above or below 30	Radial Pulse Present / Cap refill < 2 sec?	Obeys Commands?
38	n	y	Below	y	n
39	y	y	Below	y	y
40	n	y	Below	y	y
41	n	y	Below	y	y
42	n	y	Below	y	y
43	n	y	Below	y	y
44	y	y	Below	y	y
45	n	y	Below	n	y
46	n	y	Below	y	y
47	y	y	Below	y	y
48	n	y	Below	y	n
49	y	y	Below	y	y
50	y	y	Below	y	y
51	n	n	Below	n	n
52	y	y	Below	y	y
53	n	y	Below	y	n
54	y	y	Below	y	y
55	y	y	Below	y	y
56	n	y	Below	y	n
57	n	y	Below	y	n
58	n	n	Below	n	n
59	y	y	Below	y	y
60	n	y	Below	y	y
61	y	y	Below	y	n
62	y	y	Below	y	y
63	n	y	Below	y	n
64	y	y	Below	y	y
65	n	y	Below	y	y
66	n	y	Below	y	n
67	n	y	Below	n	n
68	n	n	Below	n	n
69	n	n	Below	n	n
70	n	y	Below	y	y
71	n	n	Below	n	n
72	n	y	Below	y	y
73	n	y	Below	y	n
74	y	y	Below	y	n

Patient number	Ambulatory?	Spontaneous Breathing / airway repositioning successful?	Respiratory Rate above or below 30	Radial Pulse Present / Cap refill < 2 sec?	Obeys Commands?
75	y	y	Below	y	y
76	y	y	Below	y	y
77	y	y	Below	y	y
78	n	y	Below	y	y
79	n	y	Below	y	y
80	n	y	Below	y	y
81	n	y	Below	y	y
82	y	y	Below	y	y
83	n	n	Below	n	n
84	n	n	Below	y	n
85	y	y	Below	y	y
86	y	y	Below	y	y
87	n	y	Below	y	y
88	n	y	Below	y	n
89	n	y	Below	n	y
90	n	n	Below	n	n
91	n	n	Below	n	n
92	n	y	Below	y	n
93	n	y	Below	y	n
94	y	y	Below	y	y
95	y	y	Below	y	y
96	n	n	Below	n	n
97	y	y	Below	y	y
98	n	n	Below	n	n
99	y	y	Below	y	y
100	y	y	Below	y	y
101	n	y	Below	y	n
102	y	y	Below	y	y
103	y	y	Below	y	y
104	n	y	Below	y	y
105	n	y	Below	y	y
106	y	y	Below	y	y
107	n	y	Below	y	n
108	n	y	Below	y	y
109	n	n	Below	n	n
110	n	n	Below	n	n
111	n	y	Below	y	n

Patient number	Ambulatory?	Spontaneous Breathing / airway repositioning successful?	Respiratory Rate above or below 30	Radial Pulse Present / Cap refill < 2 sec?	Obeys Commands?
112	n	y	Below	y	y
113	n	y	Below	y	n
114	y	y	Below	y	y
115	n	n	Below	y	n
116	y	y	Below	y	y
117	n	y	Below	y	n
118	n	y	Below	y	y
119	y	y	Below	y	y
120	y	y	Below	y	y
121	n	y	Below	y	y
122	y	y	Below	y	y
123	n	y	Below	y	y
124	n	y	Below	y	n
125	n	y	Below	y	n
126	n	y	Below	y	y
127	n	n	Below	y	n
128	y	y	Below	y	y
129	y	y	Below	y	y
130	n	y	Below	y	y
131	n	n	Below	n	n
132	y	y	Below	y	n
133	n	n	Below	n	n
134	y	y	Below	y	y
135	n	n	Below	y	n
136	n	n	Below	n	n
137	n	y	Below	y	y
138	n	y	Below	y	y
139	y	y	Below	y	y
140	n	y	Below	y	y
141	n	y	Below	y	n
142	y	y	Below	y	y
143	y	y	Below	y	y
144	y	y	Below	y	y
145	n	y	Below	n	n
146	n	y	Below	y	n
147	n	y	Below	y	n
148	n	y	Below	y	n

Patient number	Ambulatory?	Spontaneous Breathing / airway repositioning successful?	Respiratory Rate above or below 30	Radial Pulse Present / Cap refill < 2 sec?	Obeys Commands?
149	n	y	Below	y	y
150	n	y	Below	y	n
151	n	y	Below	y	y
152	y	y	Below	y	y
153	n	y	Above	y	y
154	y	y	Below	y	y
155	y	y	Below	y	y
156	y	y	Below	y	y
157	n	y	Below	y	y
158	n	y	Below	y	y
159	n	n	Below	n	y
160	n	n	Below	n	n
161	n	y	Below	y	n
162	n	y	Below	n	n
163	n	y	Below	n	n
164	n	y	Below	y	n
165	n	n	Below	n	n
166	n	y	Below	y	n
167	n	y	Below	y	y
168	n	y	Below	y	y
169	y	y	Below	y	y
170	n	y	Below	y	y
171	y	y	Below	y	y
172	n	y	Below	y	y
173	n	y	Below	y	y
174	y	y	Below	y	y
175	n	y	Below	y	n
176	y	y	Above	y	y
177	n	y	Below	y	n
178	y	y	Below	y	y
179	y	y	Below	y	y
180	y	y	Below	y	y

Appendix B

Thank you for choosing to help me complete this study project. This study will be assessing the accuracy of three different triage systems at placing individual patients into the correct triage category, based on the reference of the American College of Surgeons' Injury Severity Score. Your assistance will prove invaluable in eliminating the possibility of bias from my study. In order to ensure that my study is not considered to be research on human subjects, with all the regulatory concerns that entails, I will not be collecting any information about the volunteers who choose to participate. You will not be individually recognized in the final paper, but your efforts will be acknowledged collectively.

SALT Triage Instructions

You have been provided with a copy of the SALT triage algorithm, these instructions, and a list of information about patients. The list of patients will provide you only with the specific information to answer the questions asked by the triage algorithm. Your task is to follow the algorithm until reaching a category assignment, and note the category in the block provided for each patient. Every patient entry will contain the answers to ALL of the questions asked by the algorithm, not only the ones needed to assign the triage category. Once you have arrived at a category assignment, disregard any further information in that patient's entry, and do not continue further on the algorithm with that patient. Simply annotate the first triage category that the algorithm leads you to.

There are only two points for you to be aware of:

-The SALT algorithm first calls for a global sorting of all patients. This does not impact how each patient is to be assessed, only the order in which they would be assessed. This step is being omitted, as it has no impact on the final triage category

-Note that the algorithm unfortunately uses a double negative in the assessment of a patient, asking if the patient is "Not in respiratory distress?" The answer for this column is consistent with the wording on the algorithm, i.e. an entry of "y" in this field indicates that the patient is NOT in respiratory distress, and an entry of "n" indicates that the patient IS in respiratory distress. As the answers on the patient list match the algorithm's wording, follow the algorithm exactly as written.

Patient number	Breathing?	Obeys Commands /purposeful movements?	Peripheral pulse present?	Not in Respiratory Distress?	Major Hemorrhage Controlled?	Minor Injuries Only?	Likely to survive given current resources?
1	y	y	y	y	y	n	y
2	y	y	y	y	y	n	y
3	y	y	y	n	y	n	y
4	y	n	y	y	y	n	y
5	y	y	y	y	y	y	y
6	y	n	y	n	y	n	y
7	y	y	y	y	y	y	y
8	y	y	y	n	y	n	y
9	y	y	y	y	y	y	y
10	y	y	n	y	y	n	y
11	y	n	y	n	y	n	y
12	y	y	y	y	y	n	y
13	y	n	y	n	y	n	y
14	y	y	y	y	y	n	y
15	y	n	y	y	y	n	y
16	y	y	y	n	y	n	y
17	y	y	n	y	y	n	y
18	y	y	y	y	y	n	y
19	y	y	y	y	y	y	y
20	y	y	y	y	y	n	y
21	y	y	y	n	y	n	y
22	y	y	y	y	y	n	y
23	y	y	y	y	y	n	y
24	y	y	y	y	y	n	y
25	y	y	y	y	y	n	y
26	y	n	y	y	y	n	y
27	n	n	y	n	y	n	n
28	y	y	y	n	y	n	y
29	y	n	n	n	n	n	y
30	n	n	y	n	y	n	y
31	y	y	y	y	y	n	y
32	y	y	y	y	y	n	y
33	n	n	n	n	n	n	n
34	n	n	n	n	n	n	n
35	y	n	y	y	y	n	y

Patient number	Breathing?	Obeys Commands /purposeful movements?	Peripheral pulse present?	Not in Respiratory Distress?	Major Hemorrhage Controlled?	Minor Injuries Only?	Likely to survive given current resources?
36	y	n	y	n	y	n	y
37	y	y	y	y	y	n	y
38	y	n	y	n	y	n	y
39	y	y	y	y	y	y	y
40	y	y	y	y	y	n	y
41	y	y	y	y	y	n	y
42	y	y	y	y	y	n	y
43	y	y	y	y	y	n	y
44	y	y	y	y	y	n	y
45	y	y	n	y	n	n	y
46	y	y	y	n	y	n	y
47	y	y	y	y	y	y	y
48	y	n	y	n	y	n	y
49	y	y	y	y	y	n	y
50	y	y	y	y	y	n	y
51	n	n	n	n	n	n	n
52	y	y	y	y	y	y	y
53	y	n	y	y	y	n	y
54	y	y	y	y	y	n	y
55	y	y	y	n	y	y	y
56	y	n	y	y	y	n	y
57	y	n	y	n	y	n	y
58	n	n	n	n	n	n	n
59	y	y	y	y	y	n	y
60	y	y	y	y	y	n	y
61	y	y	y	y	y	n	y
62	y	y	y	y	y	n	y
63	y	n	y	n	y	n	y
64	y	y	y	y	y	y	y
65	y	y	y	y	y	n	y
66	y	n	y	n	y	n	y
67	y	n	n	n	n	n	y
68	n	n	n	n	n	n	n
69	n	n	n	n	n	n	n
70	y	y	y	y	y	y	y

Patient number	Breathing?	Obeys Commands /purposeful movements?	Peripheral pulse present?	Not in Respiratory Distress?	Major Hemorrhage Controlled?	Minor Injuries Only?	Likely to survive given current resources?
71	n	n	n	n	n	n	n
72	y	y	y	y	y	n	y
73	y	y	y	y	y	n	y
74	y	y	y	n	y	n	y
75	y	y	y	n	y	n	y
76	y	y	y	y	y	y	y
78	y	y	y	y	y	n	y
79	y	y	y	y	y	n	y
80	y	y	y	y	y	y	y
81	y	y	y	n	y	n	y
82	y	y	y	y	y	n	y
83	n	n	n	n	n	n	n
84	n	n	y	n	y	n	y
85	y	y	y	y	y	n	y
86	y	y	y	y	y	n	y
87	y	y	y	y	y	y	y
88	y	y	y	y	y	n	y
89	y	y	n	y	n	n	y
90	n	n	n	n	n	n	n
91	n	n	n	n	n	n	n
92	y	n	y	y	y	n	y
93	y	n	y	n	y	n	y
94	y	y	y	y	y	n	y
95	y	y	y	y	y	n	y
96	n	n	n	n	n	n	n
97	y	y	y	y	y	n	y
98	n	n	n	n	n	n	n
99	y	y	y	y	y	y	y
100	y	y	y	y	y	n	y
101	y	n	y	n	y	n	y
102	y	y	y	y	y	y	y
103	y	y	y	y	y	y	y
104	y	y	y	n	y	n	y
105	y	y	y	y	y	y	y
106	y	y	y	y	y	y	y

Patient number	Breathing?	Obeys Commands /purposeful movements?	Peripheral pulse present?	Not in Respiratory Distress?	Major Hemorrhage Controlled?	Minor Injuries Only?	Likely to survive given current resources?
107	y	y	y	y	y	n	y
108	y	y	y	y	y	y	y
109	n	n	n	n	n	n	n
110	n	n	n	n	n	n	n
111	y	y	y	n	y	y	y
112	y	y	y	y	y	n	y
113	y	n	y	n	y	n	y
114	y	y	y	y	y	y	y
115	n	n	y	n	y	n	y
116	y	y	y	y	y	y	y
117	y	y	y	n	y	n	y
118	y	y	y	y	y	n	y
119	y	y	y	y	y	n	y
120	y	y	y	y	y	y	y
121	y	y	y	y	n	n	y
122	y	y	y	n	y	n	y
123	y	y	y	y	y	y	y
124	y	y	y	n	y	n	y
125	y	y	y	y	y	n	y
126	y	y	y	y	y	y	y
127	n	n	y	n	y	n	n
128	y	y	y	y	y	y	y
129	y	y	y	y	y	n	y
130	y	y	y	y	y	y	y
131	n	n	n	n	n	n	n
132	y	n	y	y	y	y	y
133	n	n	n	n	n	n	n
134	y	y	y	y	y	y	y
135	n	n	y	n	y	n	n
136	n	n	n	n	n	n	n
137	y	y	y	y	y	n	y
138	y	y	y	y	y	y	y
139	y	y	y	y	y	n	y
140	y	y	y	y	y	y	y
141	y	n	y	n	y	n	y

Patient number	Breathing?	Obeys Commands /purposeful movements?	Peripheral pulse present?	Not in Respiratory Distress?	Major Hemorrhage Controlled?	Minor Injuries Only?	Likely to survive given current resources?
142	y	y	y	y	y	n	y
143	y	y	y	y	y	y	y
144	y	y	y	y	y	y	y
145	y	n	n	n	n	n	y
146	y	n	y	n	y	n	y
147	y	n	y	n	y	n	y
148	y	n	y	y	y	n	y
149	y	y	y	n	y	n	y
150	y	y	y	y	y	n	y
151	y	y	y	y	y	y	y
152	y	y	y	y	y	y	y
153	y	y	y	n	y	n	y
154	y	y	y	y	y	y	y
155	y	y	y	y	y	y	y
156	y	y	y	y	y	y	y
157	y	y	y	y	y	n	y
158	y	y	y	n	y	y	y
159	n	n	n	n	n	n	y
160	n	n	n	n	n	n	y
161	y	n	y	n	y	n	y
162	y	n	n	n	n	n	y
163	n	n	n	n	n	n	n
164	y	n	y	n	y	n	y
165	n	n	n	n	n	n	y
166	y	n	y	y	y	n	y
167	y	y	y	n	y	n	y
168	y	y	y	y	y	n	y
169	y	y	y	y	y	n	y
170	y	y	y	y	y	n	y
171	y	y	y	y	y	y	y
172	y	y	y	y	y	n	y
173	y	y	y	y	y	y	y
174	y	y	y	y	y	y	y
175	y	n	y	y	y	n	y
176	y	y	y	n	y	n	y

Patient number	Breathing?	Obeys Commands /purposeful movements?	Peripheral pulse present?	Not in Respiratory Distress?	Major Hemorrhage Controlled?	Minor Injuries Only?	Likely to survive given current resources?
177	y	y	y	y	y	y	y
178	y	y	y	y	y	y	y
179	y	y	y	y	y	y	y
180	y	y	y	y	y	y	y

Appendix C

Thank you for choosing to help me complete this study project. This study will be assessing the accuracy of three different triage systems at placing individual patients into the correct triage category, based on the reference of the American College of Surgeons' Injury Severity Score. Your assistance will prove invaluable in eliminating the possibility of bias from my study. In order to ensure that my study is not considered to be research on human subjects, with all the regulatory concerns that entails, I will not be collecting any information about the volunteers who choose to participate. You will not be individually recognized in the final paper, but your efforts will be acknowledged collectively.

ESI Triage Instructions

You have been provided with a copy of the ESI triage algorithm, these instructions, and a list of information about patients. The list of patients will provide you only with the specific information to answer the questions asked by the triage algorithm. Your task is to follow the algorithm until reaching a category assignment, and note the category in the block provided for each patient. Every patient entry will contain the answers to ALL of the questions asked by the algorithm, not only the ones needed to assign the triage category. Once you have arrived at a category assignment, disregard any further information in that patient's entry, and do not continue further on the algorithm with that patient. Simply annotate the first triage category that the algorithm leads you to.

There are two points for you to be aware of:

-There are a number of patients on the list with fields that are greyed out and contain no data. This is intentional. These patients are not suitable for use in evaluating the ESI algorithm specifically. Simply skip these entries and proceed with the next available patient.

-Final categories "3", "4", and "5" are grouped together for the purpose of this study. The patient list does not provide an answer for the number of resources needed for each patient. If the patient arrives at this question, simply move on to the "danger zone vitals?" question. If the answer to the "danger zone vitals?" question is yes, assign the patient to category "2".

Patient number	Requires immediate life saving intervention?	High Risk Situation?	Confused/ lethargic /Disoriented?	Severe Pain / Distress?	Danger Zone Vitals?
1	n	n	n	y	n
2	n	n	n	y	y
3	n	y	n	y	y
4	y	y	y	n	n
5	n	n	n	n	y
6	n	y	y	y	y
7	n	n	n	n	n
8	y	y	n	y	y
9	n	n	y	n	n
12	n	n	n	n	n
13	y	y	y	y	y
14	n	n	n	y	y
15	y	y	y	y	y
16	n	y	n	y	y
17	n	n	n	n	n
18	n	n	n	y	n
19	n	n	n	n	y
20	n	n	n	y	y
21	n	n	n	n	n
22	n	n	n	n	n
23	n	n	n	n	y
24	n	y	n	n	n
25	n	y	n	n	n
26	y	y	y	y	n
28	n	y	y	y	y
29	y	y	y	y	y
31	n	y	n	y	y
32	n	y	n	y	y
35	y	y	y	y	n

Patient number	Requires immediate life saving intervention?	High Risk Situation?	Confused/ lethargic /Disoriented?	Severe Pain / Distress?	Danger Zone Vitals?
36	y	y	y	n	y
37	n	n	n	y	y
38	y	y	y	y	n
39	n	n	n	y	y
40	n	y	n	y	y
41	n	y	n	y	y
42	n	n	n	y	y
43	n	n	n	y	y
44	n	y	n	n	n
45	y	y	n	n	n
46	n	y	n	y	y
47	n	n	n	n	n
48	y	y	y	y	y
49	n	y	n	n	n
50	n	y	n	y	n
52	n	n	n	n	n
54	n	y	n	n	y
55	n	n	y	y	y
56	y	y	y	n	n
57	y	y	y	y	y
59	n	n	n	n	n
60	n	y	n	y	n
61	n	y	y	y	y
62	n	y	n	y	y
63	y	y	y	n	n
64	y	n	y	n	n
65	n	y	y	n	n
66	y	y	y	y	y
67	y	y	y	y	n
70	n	y	y	y	n

Patient number	Requires immediate life saving intervention?	High Risk Situation?	Confused/ lethargic /Disoriented?	Severe Pain / Distress?	Danger Zone Vitals?
72	n	n	n	n	n
73	y	y	y	y	y
74	y	y	y	y	y
76	n	n	n	n	n
77	n	n	n	n	n
78	n	y	n	n	n
79	n	y	y	y	y
80	n	n	n	n	n
81	n	n	y	y	y
82	n	y	n	n	n
85	n	y	y	n	n
86	n	y	n	n	n
87	n	y	n	n	n
88	n	y	y	n	n
89	y	y	n	y	y
94	n	n	n	n	n
95	n	n	n	n	n
97	n	n	y	n	n
99	n	n	y	y	y
100	n	n	n	n	n
101	y	y	y	y	y
102	n	n	y	n	n
103	n	n	y	n	n
105	n	n	n	n	n

Patient number	Requires immediate life saving intervention?	High Risk Situation?	Confused/ lethargic /Disoriented?	Severe Pain / Distress?	Danger Zone Vitals?
106	n	n	n	n	n
108	n	n	n	y	n
111	n	y	y	y	y
112	n	n	n	n	n
113	y	y	y	y	y
114	n	n	n	n	y
116	n	n	n	n	n
117	n	y	y	y	y
118	n	y	n	n	y
119	n	n	n	n	n
120	n	n	n	y	y
122	n	n	n	y	y
123	n	n	n	n	n
124	y	y	y	n	n
126	n	n	n	n	n
128	n	n	n	y	y
129	n	n	n	y	y
130	n	n	n	n	n
132	y	y	y	n	n
133	y	y	y	n	n
134	n	n	n	n	n
138	n	n	n	n	n
139	n	n	n	n	y
140	n	y	n	n	n

Patient number	Requires immediate life saving intervention?	High Risk Situation?	Confused/ lethargic /Disoriented?	Severe Pain / Distress?	Danger Zone Vitals?
142	n	n	y	n	n
143	n	n	n	n	n
144	n	n	n	y	y
146	y	y	y	y	y
147	y	y	y	y	y
149	n	y	n	y	n
151	n	y	n	n	n
152	n	n	n	y	y
154	n	n	n	n	n
155	n	n	n	n	n
156	n	n	n	y	y
157	n	n	n	n	n
158	n	n	n	y	y
161	y	y	y	y	y
165	y	y	y	y	n
168	n	y	y	n	n
169	n	n	n	n	n
170	n	y	n	y	y
171	n	y	n	n	n
172	n	y	n	n	n
173	n	n	n	y	y
174	n	n	n	n	n
175	y	y	y	n	n

Patient number	Requires immediate life saving intervention?	High Risk Situation?	Confused/ lethargic /Disoriented?	Severe Pain / Distress?	Danger Zone Vitals?
176	n	y	y	y	y
177	y	y	y	n	n
178	n	n	y	n	n
179	n	n	n	n	n
180	n	n	n	n	n